

What comes after the INF treaty?

Getting from last week's treaty on intermediate missiles to the next agreement on strategic missiles will not be easy. Negotiators should cut their teeth on a surveillance negotiation.

Of the many things to be said about the US-Soviet treaty on nuclear missiles of intermediate range (INF), signed last week in Washington, the most certain is that it will restore the strategic balance in Europe to roughly what it was in the mid-1970s. It was then that Soviet SS-20 missiles first appeared and the North Atlantic Treaty Organization decided (exactly eight years ago) that its countervailing forces should be "modernized". The return to the *status quo* is much to be grateful for: Europe has been a more turbulent, although not necessarily a more dangerous, place since the arrival of the intermediate-range missiles. But although the improvement of military technology and progress towards arms control are antithetical processes, they do not exactly cancel out. That is one reason why it is difficult to tell what the future will be like.

Much will depend on the degree to which the two signatories have accepted that the INF treaty can be but a step (not the first) in a longer and more ambitious process. On the face of things, the signs are good. The chief substance of last week's negotiations in Washington seems to have been the most detailed discussion yet of the proposed 50 per cent cut in the two superpowers' stock of strategic missiles. Both President Reagan and Mr Mikhail Gorbachev say that this new treaty commands the centre of their attention; they say they hope to have a treaty they can sign in the first half of 1988. The enthusiasm is welcome; let us hope the optimism is well grounded.

There are two obvious obstacles to the search for a strategic agreement not intrinsic to the project itself. First, although the US Senate (which must ratify the INF treaty) will probably give its consent to the treaty now signed (and there are ample precedents for the INF treaty sticking even if the Senate makes trouble), the debate due to begin next month may be a bruising process. Second, there is the US Strategic Defense Initiative (SDI), the stumbling block at Reykjavik earlier in the year which is said also to have been part of the reason for last week's slow progress. That is why it is important that neither side should pin all its hopes on a quick strategic arms agreement, natural successor to INF though it may be.

Violation

Both obstacles must be recognized for what they are if they are to be surmounted successfully. The objections to the INF treaty likely to surface in the US Senate have been well flagged and are of two kinds — technical and political/strategic. Although the new treaty differs from all predecessors (apart from the unratified treaty on the Peaceful Uses of Nuclear Weapons) in providing for both regular and challenge on-site inspection, there will be complaints that it does not reduce to zero the chance of undetected violation.

That is an unremarkable truism. No treaty can do that. The principle is merely that the chance that violations of any kind will be detected is high enough to deter those inclined to cheat. In that sense, last week's treaty is unprecedented in the detail with which verification measures have been spelled out. Senators inclined to complain that there are still gaps would be better advised to worry that the treaty provisions are so complicated that they they will certainly prove to be ambiguous on technical

issues not so far anticipated. That is exactly what has happened with the Anti-Ballistic Missile (ABM) Treaty, where the discordance between the body of the treaty and a later protocol stems from a failure of the drafters to anticipate what is now technically possible in spaceflight.

SDI is evidently a more serious matter. In logic, President Reagan is right to say that a defensive shield capable of destroying all hostile missiles would be a boon, rendering strategic missiles impotent (and thus useless) and even making many arms control negotiations pointless. Troubles arise chiefly because no shield of this kind can be perfect, while there are solid grounds for the widespread scepticism that it could be nearly so. (The difficulty of arranging defences against missiles launched from submarines should be a sufficient proof of that.) In the circumstances, some future deployment of a partially effective SDI would have the effect of persuading the other side to deploy more strategic missiles. That is why, if the United States and the Soviet Union do succeed in reaching an agreement on strategic missiles, in equity the side retaining the right to deploy SDI should be prepared to let the other keep a larger missile striking force (see *Nature* 330, 96; 1985).

Undermined

Some in the US Senate will say that these considerations have been undermined by Mr Gorbachev's admission, on the eve of the talks in Washington, that the Soviet Union is also engaged on research in ballistic missile defensive systems. But that does not give the US administration a free hand. Under the ABM treaty, the Soviet Union has kept the defensive system it installed around Moscow in the early 1970s; is it likely that the Soviet Union would have neglected the possibility of improving that, or that it would have neglected to follow the United States in the exploration of the well-advertised components of SDI? What matters about SDI is not research, but deployment.

But what does deployment mean? Enthusiasts for SDI suppose it to be a unitary process. On the analogy of an umbrella, one has merely to flick the catch and hoist the canopy. But even a decade from now, the launching of the few hundreds of (still untested) satellites required for a putatively effective shield will not be a simple matter. Present launching capacity in the United States would hardly be able to keep pace with natural failure rates. But on its own admission, the SDI project has still not decided what systems it would use for destroying hostile missiles in different circumstances. Even if the claim that SDI would ultimately be an effective umbrella were valid, its several components will become usable only at different times.

That is why President Reagan's assurance in his weekend radio broadcast that nothing agreed last week will inhibit a decision to "deploy" SDI is over-simple, to say the least. It also appears innocent of his earlier undertaking to the British Prime Minister, Mrs Margaret Thatcher, which has since been repeated publicly, that there will be negotiations with the Soviet Union before SDI is deployed and, gratuitously, that the United States will also be willing to discuss "sharing" the results of SDI research.

Those looking for constructive ways out of the SDI impasse



should begin at this point, by asking what is meant by the sharing of research results. It is not necessary to wait for SDI to be complete to make a start. The place at which to begin is that essential component of SDI, as described in the United States — a system of satellites carrying infrared sensors to detect missiles in the early stages of their flight, but surveillance satellites, also important sources of data for the verification of last week's treaty, will be essential for monitoring performance under its strategic successor. So why not begin right away, by negotiating an agreement on "sharing" of surveillance technology, present and future? To those most concerned, this may seem an outrageous suggestion, given that the resolution achieved by surveillance satellites is one of the most closely guarded of military secrets. "We do not want the other fellow to know what we can see" is the usual argument. But that position must now have been seriously undermined by the mere exchange of information required for last week's treaty. Moreover, the present regime will be the more stable if each party to the INF treaty knows that the other is well-equipped to know that it is keeping to the rules. That is one reason why an early negotiation on surveillance would be a useful prelude to what must follow; even a failure to make progress would be a useful warning of troubles that may lie ahead.

More generally, the undertaking that negotiations will precede deployment is politically important. It is not merely that President Reagan may have more to fear from Mrs Thatcher's tongue than even from Mr Gorbachev's, but that Western European governments are likely to regard compliance with the undertaking as an important determinant of their future relationships with the major powers. Senators who hold that Western Europe would be better able to make that demand if it shouldered more responsibility for its own defence will be missing the important point. (But the INF treaty will stimulate close attention to the modernization of conventional forces, while this week's discussions in London between British and French ministers on the joint development of air-launched missiles is another sign of the way the wind is blowing.) The general enthusiasm for last week's treaty should not conceal the diversity of the reasons for rejoicing.

That the parties to arms control agreements (five Western European governments have formally accepted that the Soviet Union has a right to monitor the removal of US missiles) may have different motives is by no means unusual. Especially to mainland Europeans, one of the still-distant prizes of the arms control process is that Central Europe may thereby become more like the politically free and culturally rich place it used to be, but without the recurrent conflicts that made Europe unsafe as well. It seems now to be generally understood, perhaps in the East as well, that this happy state of affairs will not be possible until the issues that divide people are the sufficiently contentious quarrels that arise over trade, broadcasting and competitive sport. Sensible Europeans acknowledge that there is a long way to go before that will be the case, but they also fear that super-power tension over the deployment of SDI could make their legitimate goal less easily attainable.

None of this is meant to rubbish last week's negotiations. Both the major powers have earned credit by their demeanour and even by their evident realism about the difficulties they face. But the danger is that they may not fully appreciate that the effect of INF is not simply to put the clock back to say 1975, before *détente* began to unravel. Western Europe is stronger economically, yet Japan (spending one per cent of GNP on defence) is the engine of the world's economy. By the same tests, both the major powers have faltered, the United States only in a relative sense but the Soviet Union to such a degree that the state of the Soviet economy appears to be the mainspring of the new drive for change, almost without regard to where it may lead. Two lessons can be drawn from that. First, history will not of its own accord repeat itself, meaning that the strategic agreement will not follow INF as, in the 1970s, SALT II followed SALT I.

Second, the economic benefits of military restraint, long understood by middle and small powers, have now become respectable prizes for the superpowers. It will be interesting to see how quickly President Reagan follows Mr Gorbachev in heeding that truth. □

Who pays for health?

British health services are heading for another crisis; they need a public policy.

THE British National Health Service, known in some quarters as "socialized medicine", is a splendid institution once again in trouble. For the past several months, the newspapers have been full of tales of how urgently necessary surgery has been postponed for lack of skilled people, or of the funds with which to employ them. Last week, the presidents of the three principal colleges of medicine (physicians, surgeons and obstetricians) startled the government (and probably some of their own members as well) by issuing a public statement that the Health Service is urgently in need of more resources or in danger of collapse. The government's reply is that spending on the health services is indeed increasing, by roughly £1,000 million a year, and will continue to increase for the foreseeable future. Yet so will the gap between expectation and reality.

The nature of the problem is familiar, not only in Britain (West Germany and the United States have institutionalized different resolutions of the same dilemma) nor only in the provision of health services (British education has been squeezed for similar reasons). Public services, as the British National Health Service remains, are notoriously always under financial pressure from rising expectations. Health services are especially vulnerable, for expectations rise to encompass immortality. The British problem is especially acute because the national wealth has been growing less quickly than the standard of medical treatment elsewhere, not to mention spending; US health costs are now nearly 11 per cent of gross national product (GNP), compared with 8 per cent of a smaller (GNP) in Britain (amounting to roughly £400 per head a year). The present government, while committed to the maintenance of the National Health Service (if not from conviction, then for sound political reasons, the wish not to offend too many voters), seems to be hoping for some kind of undefined collaboration with the private sector, largely that supported by medical insurance schemes.

That is potentially too risky a way of going forward. While there are few other countries in which private medicine is entirely out of court (even the Soviet Union has a little) there appears to be none in which it is in roughly equipotent balance with the public services. Nor is stability likely in a thoroughly mixed system. To the extent that insurance-based medicine may be financed at somewhat arbitrary levels, private medicine is likely always to be able to draw scarce resources, mostly people, from the public sector. In Britain, that appears to account for the present shortage of specialized nurses in the public sector. The real danger is that a continuation of this trend could hamper the capacity of the public service to perform its notable civilizing services — the care of the old, the young, the indigent and the chronically sick.

That is why almost any policy would be preferable to the present pragmatism. The government's encouragement of private medical insurance (premiums nevertheless do not earn tax relief, as in the United States) may ease its budget problem, but in the long run will increase unit costs for medical services. If that is the route that is to be followed, compulsory medical insurance (as in West Germany) for those at work would be preferable. But in the short run, there seems no alternative to extra public spending. The British government, with a budget only a few months away, would prefer to cut taxes, which might well be the less popular course. □

US/Soviet science summit parlay in new spirit

- Upbeat mood prevails in Washington
- Economic changes forecast

Washington

THE optimism and enthusiasm at last week's summit meeting here between General-Secretary Mikhail Gorbachev and President Ronald Reagan spilled over into science during a nearly simultaneous meeting between members of the Soviet and US academies of science.

Frank Press, president of the US academy, described the meeting as an object lesson in *glasnost*. "The Americans who attended the meeting had all read about *glasnost*", said Press. "But that did not prepare us for what we encountered."

The gathering at National Academy of Sciences headquarters across the street from the State Department was billed as a meeting on issues of science, technology and economic development. Some three dozen American and one dozen Soviet scientists, economists and corporate executives sat around a large table, as translators struggled to keep pace with the often animated discussion.

Economic issues dominated the meeting. Abel Aganbegyan, secretary of the economics department of the Soviet academy, described the changes as a shift from extensive to intensive development. In practical terms, this will mean spending more money on factory machinery, less on new buildings. It will also mean giving more freedom to individual enterprises to chart their own course, taking a cue from the demands of their customers. Prices for goods will be allowed to fluctuate in response to demand. The stratified and centralized agencies that make economic decisions will be made smaller and more responsive to local needs.

But with *perestroika* — the reconstruction phase that follows *glasnost* — there will also be upheaval. Restructuring entrenched bureaucracies will mean constant political battles. Installing new technology inevitably means a delay in production until the new machinery is brought on line.

The Soviet delegation was headed by Yevgeniy Velikhov, vice-president of the Soviet academy, who sounded a familiar theme at the start of the meeting: the need for bringing scientific and technological achievements to industry. He also spoke of education reform. At the university level, there will be academic chairs at research facilities so that teaching can go on at places traditionally reserved for research only. Likewise, Velikhov says more money will be spent on basic

research at universities.

At the high-school level, the Soviet Union plans to have one million Soviet-built personal computers in the classroom by 1992. That forecast brought a warning from John Scully, chief executive officer of Apple Computers, who told Velikhov that scaling up production of computers places tremendous demands on quality control. He also predicted that hardware development would continue to proceed rapidly, ultimately leading to desk-top supercomputing within the decade. Scully urged Velikhov to steer Soviet participation in computers in the direction of software development to anticipate the demands of the new hardware.

The meeting yielded no concrete proposals. Roald Sagdeev, director of the Soviet Space Research Institute, described the need for developing standards for data exchange as international activities begin on the international geosphere/biosphere programme. US Environmental Protection Agency administrator Lee Thomas urged the Soviet Union to support the recently concluded agreement on controls on chlorofluorocarbons to protect atmospheric ozone. A discussion on AIDS was postponed when the Soviet delegation had to leave for the White House to attend the signing ceremony for the intermediate nuclear force treaty.

Despite the lack of substantive exchanges, the mood was clearly different from previous US/Soviet discussions. "The evasiveness and rehearsed responses of the past were gone, replaced for the most part by openness and frankness", said Press. But if openness characterized most of the meeting, such was not the case for the issue of human rights. Vladimir Kudryavtsev, director of the Institute of State and Law described claims of large numbers of potential Jewish emigrés as "fantastical propaganda".

The decorum of the meeting broke down only when an aide to Press informed him that Raisa Gorbachev would soon be outside, to see the bust of Albert Einstein in front of the building. A mass exodus from the room ensued, on the strength of what turned out to be a false alarm, and the meeting resumed. There was a second exodus when the aide reappeared, but Mrs. Gorbachev never left her limousine. In the uncertainty, it fell to Sagdeev to propose the obvious; continue the meeting.

Joseph Palca

All systems go for star wars despite summit

Washington

JUST a few days after the end of his summit meeting with Soviet leader Mikhail Gorbachev, US President Ronald Reagan has reiterated his determination to continue with the Strategic Defense Initiative (SDI). In a weekend radio address, he described the building of a nuclear missile defence as a "moral imperative" and insisted that he would not give up his plans in order to achieve the 50 per cent cut in nuclear forces aimed for in the Strategic Arms Reduction Talks (START).

Disagreements between the two sides over SDI continued until the last minutes of the summit and the final compromise statement achieved little of what the Soviets had wanted. They had proposed that SDI testing be limited by the Anti-Ballistic Missile (ABM) Treaty as "signed and ratified", the word "ratified" implying that the view of the US Senate, that testing in space of a missile defence system would violate the treaty, must be upheld. But "ratified" was removed after US protests, leaving Reagan free to fight for his non-restrictive view of the treaty.

Despite continuing Soviet condemnation of SDI, there were hints that the problem would be put aside to allow progress in the START negotiations.

Although the two sides agreed to reduce their nuclear arsenals to 4,900 ballistic missile warheads each, the sub-limits on the different warheads remain unclear, as do restrictions on sea-launched cruise missiles and the new multiple-warhead missiles on US Trident submarines.

Not everything went the US administration's way over SDI. The White House wanted a statement allowing the United States to withdraw from the ABM treaty if it became necessary to do so in order to carry out testing of SDI in space. But the final communique denied permission to withdraw from the treaty.

In the short term, Congress seems more likely to inhibit the development of anti-missile systems than does Soviet opposition. In the defence authorization bill, agreement has been reached on a total budget of \$3,900 million for SDI research, \$2,000 million less than the White House wanted and only 11 per cent more than last year. But attempts to force Congress's narrow interpretation of the ABM Treaty on President Reagan by including restrictions in the authorization bill were not wholly successful. The final version prevents any immediate tests that are inconsistent with the narrow interpretation of the treaty, but does not prevent planning for tests in later years.

Alun Anderson

Maintaining anencephalic babies causes consternation in USA

San Francisco

LOMA Linda University Medical Center in southern California is preparing to implement a controversial new policy on anencephalic neonates (born with only a brain stem). The hospital will allow them to be maintained on a respirator so that their organs may be used for transplantation.

By US law, organs may not be taken from a living human being until it is legally brain dead. Many anencephalics are still-born, but those born alive are allowed to die naturally. Brain-stem activity can persist for weeks or months, marked by periods of irregular breathing that deprive the organs of oxygen. This often means an anencephalic's organs have deteriorated too much to be used for transplants.

Although most adult organ donors — usually accident victims — are on life-support systems before death, the life support is meant to be life-saving. With anencephalics, some now question the morality of using life support for the sole purpose of harvesting organs.

Some physicians and lawmakers want to change the legal definition of death to define anencephalics as brain dead at birth, so that their organs can be used immediately. But that proposal has sparked heated debate, with some saying that such a move would blur the line between life and death, paving the way for the use of organs from other critically ill patients, such as those in vegetative coma.

Arthur Caplan, director of the University of Minnesota Center for Biomedical Ethics, argues that no lines are blurred with anencephalics. They are unique, not because they are dying, he says, but because they have no higher brain function: no ability to feel, think or suffer.

Last January, Caplan proposed a compromise in which anencephalic children would be kept on a respirator to preserve their organs and, when their brain-stem function ceased, be used as sources of organs. Such a protocol was used in October by physicians at the University of Western Ontario, and provided a heart for a recipient at Loma Linda hospital.

Loma Linda, a leader in the field of neonatal heart transplants, has already shown a willingness to try new approaches to resolving the acute shortage of infant organ donors. In 1984, a medical team there used a baboon heart in a human infant.

Loma Linda will be the first hospital to implement a protocol for the maintenance of anencephalics in the United States. The first infant to be kept alive in the new protocol was to have been born earlier this

week. The parents of the child learned about the fetus's condition three months ago, but chose to carry it to term in hopes of offering its organs for transplantation.

Loma Linda neonatologist Joyce Peabody, author of the hospital's new protocol, says that organ donation should be considered only when parents of anencephalic children decide on their own, without persuasion. The procedure will provide respectful treatment of the infant, including traditional provisions for comfort and pain relief, although anencephalics lack the capability to experience or interpret pain.

Because of the consensus that indefinite maintenance of an anencephalic would be

immoral, Loma Linda has adopted Caplan's suggestion that the infant be kept on a respirator for just seven days. If brain death occurs, as measured by periodic checks for cessation of spontaneous breathing, recipients will be sought for the child's heart, liver, corneas and possibly kidneys. After seven days, if the brain stem is still functioning or if organ recipients are not found, the child will be removed from the respirator.

To avoid the issue of conflict of interest, as Loma Linda is a transplant centre, Peabody said she would prefer that another hospital take on the case. However, that possibility is remote, and Loma Linda is ready to put its protocol into practice. Peabody pointed out that recipients can come from anywhere, and she hopes they will arise at sites other than Loma Linda, to simplify the hospital's role in the case.

Marcia Barinaga

Drastic changes proposed for Australian higher education

Sydney

AUSTRALIA's new Minister for Employment, Education and Training, John Dawkins, has proposed dramatic changes to Australia's higher education system. His aim is to make its institutions, some of which he describes as ossified, serve national needs "flexibly and effectively".

The essence of the change can be summed up as "increased competition". If the Dawkins plan, presented in a 126-page policy paper, wins approval, then the binary system, which separates the financing of universities and of colleges of advanced education, will be ended. All higher education institutions will have access to the same funds, while all university departments will no longer enjoy the right to resources for research.

Instead, institutions will have to bid for funds by presenting their "educational profiles" to the government, which will assess them for efficiency or effectiveness in terms of national priorities.

For funding purposes, student numbers will be assessed in terms of those who graduate rather than those who enrol. According to the policy paper, there has been an enormous waste of resources as a result of the high drop-out rates from the first years of courses.

As an initial step, it is planned that one per cent of funds in the 1988–91 triennium would be put into a pool for competitive allocation. Any institution which chose not to compete would simply lose a proportion, which will grow with time, of what it might otherwise expect.

Participating institutions, on the other hand, would also be freed from present government procedures for the approval of new courses and from rigid restrictions

on staff pay scales. Institutions would be allowed to pay academics on the basis of merit and market conditions, with the understanding that in some disciplines staff would be paid more to prevent their defection to industry. Institutions would also have to devise means whereby incompetent or redundant staff could be sacked.

To achieve economies of scale, smaller institutions will be amalgamated. The minimum size for research institutions will be set at 8,000 students. A unified system of course accreditation and a common academic year will help staff and students move between institutions.

Dawkins hopes to increase the number of graduates from 88,000 a year at present to 125,000 per year in 2001, a significant increase considering the smaller proportion of young people in the community by then. But he is in favour of those who benefit from education paying for it. Dawkins is likely to meet strong opposition to the reintroduction of fees from his Labor party colleagues who originally abolished them.

A spokesman for the association of university vice-chancellors was in favour of the removal of restrictions on institution's operations and freedom in staff remuneration, but wary of increased political intervention and, in particular, of the nature of the government's "educational profiles".

To the colleges of advanced education, the paper offers the opportunity finally to play a full role in higher education. A spokesman for the colleges' principals and directors described it as "one of the most significant junctures in the planning and provision of higher education in Australia".

Charles Morgan

University Earth scientists are preparing for reorganization

London

RESTRUCTURING an entire scientific discipline within a nation's university system is fraught with political and practical difficulties, as Britain's beleaguered geoscientists and the University Grants Committee (UGC) are discovering.

Last May, at the UGC's request, a committee under Professor E.R. Oxburgh produced a report suggesting a way of making more efficient use of the funds available for teaching and research in the Earth sciences in Britain's universities. Oxburgh's solution was to create three tiers of university department — some (about ten) equipped for research and teaching to an international standard, some equipped for the full range of teaching but with little research capability and some providing undergraduates with an understanding of the Earth sciences, probably contained within other departments, such as physics. The departments would be classified as types 1, 2 and 3, respectively.

After extensive consultation, modifications to the Oxburgh report were agreed, where two types of department would be created. Type A departments would be involved in research and honours degree teaching, and would not normally have less than the equivalent of 180 full-time students, and a staff:student ratio of about 1:9; there would probably be fewer than 20 Type A departments.

Type B departments would be contained within other departments, such as physics, and would not normally be equipped to teach single honours. Superimposed on the two-tier system would be equipment centres within a limited number of the larger departments, concentrating expensive analytical equipment. The UGC initially implied that there would be about five such centres, and that their locations would be designated before universities invited to submit bids to operate either Type A or Type B departments. The proposal was condemned by the heads of the geoscience departments, who argued that no individual department should be required to service heavy equipment across the entire range of Earth sciences and its subdisciplines, and that distribution of equipment should be decided during the bid process and not precede it. The modifications have been adopted by the UGC.

The UGC has set itself an ambitious timetable. Finalized bids must be submitted by 1 January 1988, with the new system operating by the beginning of the academic year 1988–89. One national and three regional committees have been set up to see that the whole process runs

smoothly, and an 'equipment rationalization committee' will decide where to put expensive experimental and analytical apparatus.

For the heads of departments, the task of compiling a bid is proving monumental. The UGC requires detailed completion of a 55-page questionnaire, with information ranging from the age of academic staff to five years' worth of publication and citation data. Academics are complaining that faced with such a daunting task, routine research and teaching is suffering. Privately, many feel that the bid process has been reduced to a number-shuffling exercise, with the object of arriving at the required staff:student ratio in order to qualify for Type A department status. In some cases this has meant merger negotiations with neighbouring institutions or, possibly with more damaging implications, with neighbouring departments within an institution. The difficulty arises with disciplines on the periphery of Earth sciences, such as meteorology or oceanography, that now reside within non-Earth-science departments. At the University of Newcastle upon Tyne, for example, geophysics forms part of the physics department. If it were lumped in with geology in order to produce a feasible 'Type A' bid (something that has been discussed but seems unlikely to go ahead), one of the country's largest universities would be left with an unviable physics department.

Simon Hadlington

Well done, Pioneer

Washington

PIONEER 8, designed to send back signals from interplanetary space for six months, has done quite a bit better than that.

Earlier this week it marked 20 years on the job, and is still sending back data from one of its experimental packages.

Pioneer 8 was the third in a series of four satellites launched in the 1960s to study solar flares and related phenomena. Of the eight instruments on the spacecraft, only the electric field detector is now operating.

From 1971 to 1984 the spacecraft was powered down to save the cost of monitoring the telemetry and to conserve power. But when the more sophisticated electric field detector on Pioneer 9 failed, Pioneer 8 was powered up in time to send information back when spacecraft's orbit sent it through the Earth's magnetic wake.

Also this week, Pioneer 6 celebrates its twenty-second year of operation, making it the oldest functioning spacecraft in interplanetary space.

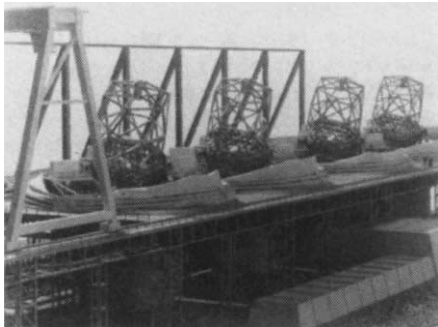
Joseph Palca

New European telescope for the Andes

Paris

THE European Southern Observatory (ESO), whose headquarters are in West Germany, has given the official go-ahead for a new telescope in northern Chile. The decision came with the announcement last week that France will release FF338 million (£33.8 million) over ten years as a contribution to the cost of building the Very Large Telescope (VLT).

West Germany has already promised an equal amount, 26.7 per cent of the total,



A model of the four connected telescopes that make up the Very Large Telescope.

and Italy a slightly smaller amount. The remainder will come from ESO's other five member states: Belgium, Denmark, the Netherlands, Sweden and Switzerland. Although Belgium is not yet committed, more than 75 per cent of the cost has been pledged, making it possible to award some of the contracts for mirror manufacturing and polishing.

VLT will eventually comprise four 8-m telescopes, giving it a power equivalent of a single 16-m telescope, greater than the US Keck telescope which will be built in Hawaii in 1990. Coupling the telescopes together to form an interferometric array involves, says Dr Couturier of Institut National des Sciences de l'Univers in Paris, "a technical gamble, but ESO is confident that it will pay off". Working in this mode, VLT will have resolution five times greater than any existing Earth-based telescope.

Two sites have been proposed for VLT — the present ESO site at La Silla at an altitude of 2,400 m, where 13 telescopes are already in operation, or a new site at Cerro Paranal, at 2,700 m. A decision will be needed by 1990 for the installation and operation of the first telescope at the end of 1993.

New techniques of mirror design and of active surface control will be needed to construct the telescopes which, once installed, will be able to operate remotely from the ESO centre at Garching, near Munich.

Peter Coles

West German health care reform hazards pharmaceutical research

Munich

THE West German pharmaceutical industry is protesting against a proposed reform in West Germany's national health insurance programme. The proposal, introduced by the government on 3 December, calls for a two-thirds reduction in the amount that public health insurers, or *Krankenkassen*, will pay for drugs.

Drug manufacturer Hoechst denounced the reform as a "death blow" to pharmaceutical research, while rival Bayer predicted that the reform would lead to the "structural collapse" of West German drug research.

Labour Minister Norbert Blüm's (Christian Democrat) reform proposal is an attempt by the ruling conservative coalition to slow galloping increases in health care costs. Under the mandatory public health insurance system, half the premiums are paid by employers and half by employees. These premiums increased in 1987 to 12.7 per cent of the average worker's income, compared with 8.2 per cent in 1970. If costs were to rise unchecked, employer and employee would be paying 20 per cent of gross income for health insurance by the year 2000.

Under the reform plan, the insurers would realize an annual saving of DM14,300 million beginning in 1989. Part of the money would derive from the reduction in drug costs and part of it from a reduction in a variety of benefits. DM6,500 million would go into a new home care programme for the severely disabled, DM500 million to the detection and prevention of cancer and other diseases, and the rest would be returned in the form of reduced premiums.

Under Blüm's plan, public insurers would pay only a certain amount for prescriptions. The amount would be fixed somewhere in the lower third of the price spread. The patient would have to pay the difference for a higher-priced drug, that is, a brand-name drug patented by a drug company, with the same effects.

Such wholesale support for generic drugs is unrealistic, the drug companies have replied, given the costs of research. Development of a single synthetic anti-

biotic cost Bayer DM330 million, said Hermann Strenger, chairman of the Bayer board. Bayer patented aspirin in the late 1890s. Blüm also called for the pharmaceutical industry to show "solidarity" by freezing prices at current levels in addition to pricing their drugs more competitively. This demand is equally galling to the drug companies, which claim to have "already done their part" by keeping prices stable over the last several years, said Helga Hennemann of Hoechst.

The drug companies complain that Blüm's reform will cost them DM3,700 million annually, a reduction of 40 per cent in the amount they normally receive from the *Krankenkassen*. Blüm places the figure at DM2,500 million.

Bayer, Hoechst and their West German rivals feel threatened by US and Japanese competition in the area of drug development. But it will be hard to assess the real effects of Blüm's reform on the companies' research budgets. At Bayer, for example, the pharmaceuticals division accounts for only 15 per cent of annual sales. The companies have until the end of 1988 to try to sway the Bundestag before Blüm's plan is made law. **Steven Dickman**

UK under fire for refusal to increase ESA science budget

London

THE British government's insistence that the science budget of the European Space Agency (ESA) be frozen after 1989 is causing growing resentment among space scientists at home and abroad. At ESA's ministerial meeting in The Hague last month (see *Nature* 330, 195; 1987), Britain was the only member state to veto previously agreed increases in the budget of the Horizon 2000 science programme. Because the science budget forms part of the mandatory subscription, its acceptance requires unanimous agreement.

This week saw the start of the familiar and usually fruitless task of persuading Britain to toe the international line, with the most vociferous campaigning coming from Britain's own space community, facing ever-increasing international isolation.

Horizon 2000, announced in 1984, was seen as a pioneering move that encouraged unprecedented unity and cooperation among Europe's space scientists. At The Hague, Britain argued that over the current five years the 27 per cent increase in real terms in the programme's budget ought not be added to, and that the programme's projected duration should instead be increased. This, argued Britain, would neither "destroy the programme nor make the intervals between missions too long".

It had originally been intended that the

1988 budget of 176.8 MAU (million accounting units; 1 AU = £0.64) be increased annually by 5 per cent to a ceiling of 232.3 MAU by 1994. Supporters say that Britain's stance will substantially damage the programme: plans had been drawn up around precisely calculated funding levels and a simple extension of the programme's life would not only be ultimately more expensive, but would disrupt the entire timetable.

A final decision on the shape of the science programme will be taken next Spring. ESA officials have until then to change the British government's mind. They are not hopeful. **Simon Hadlington**

Directorship for Roy Woodruff

San Francisco

ROY D. Woodruff, the Lawrence Livermore National Laboratory (LLNL) physicist who accused Edward Teller and Lowell Wood of misrepresenting the progress of X-ray laser research, has been appointed director of weapons verification research at LLNL.

Woodruff resigned as director of the X-ray laser programme in 1985 following his disagreement with Wood and Teller about their "overly optimistic" presentation of the programme to officials in Washington (see *Nature* 329, 751; 1987). He later filed a grievance with the University of California, which oversees the laboratory, claiming he had been demoted as punishment for his views.

In October, the university ordered the laboratory to reinstate Woodruff in a position that would make appropriate use of his skills.

George H. Miller, associate director for defence systems, said that he created Woodruff's new position of assistant associate director for verification because he expects verification to play an increasingly important role in arms control and weapons design issues, both at the laboratory and in Washington. Marcia Barinaga

Soviet psychiatry

London

THE United States Department of Health and Human Services has turned down a proposal from the Soviet Union for a joint commission of Soviet and American psychiatrists to look into allegations of the political misuse of psychiatry in the Soviet Union. Speaking on Soviet television last month, Academician Evgenii Chazov, the Soviet Minister of Health, said the head of the foreign relations administration of the US Health Department had replied that US psychiatrists are not interested in Soviet psychiatry, and would therefore not take part in any such commission. **Vera Rich**

Medical Research Council cuts inevitable in coming year

London

A 20 PER cent cut in the number of short-term grants to university scientists from the UK Medical Research Council (MRC) and a decrease in the proportion of medium-term proposals that can be supported are inevitable for the coming year. The former cut is the least damaging way to absorb the financial problems of MRC, says its new secretary (chief executive), Dr David Rees. The latter is more a reflection of a cyclical increase in demand.

Financial problems, stemming from government decisions, continue to dog the attempts of the council to counteract what Rees says is a situation where Britain is



Dr David Rees

"dangerously close to not having a critical mass in modern biology". The immediate problem is that pay awards to MRC staff continue to outstrip the growth of the council's budget. But in the foreword to the 1986-87 annual report, published this week, MRC also expresses concern over the general inadequacy of funds for medical research and warns of lost opportunities.

With little relief in sight, Rees foresees a period of more stringent evaluation of the long-term units that MRC finances, more efficiency in their operation and the evolution of new criteria by which to select what research to support in universities.

Short-term 'project grants' need particular attention, he says. Fewer must be made available to established scientists in mainstream areas who have failed to obtain more solid support; more should be available for young scientists and for specialized areas of research.

As a firm believer in a mixed 'economy' of research, Rees expects some priorities to evolve — cell biology and the human genome are examples — but is also firmly committed to the support of a "continuous infrastructure". In addition, MRC will maintain its commitment to financing clinical and health-care research on important medical problems. Scientific excellence is not the appropriate sole, or

even main, criterion for selection in this case acknowledges Rees, who will be least at home in this area as MRC's first secretary without a medical qualification. ("All else being equal", he says, "it would be better to have someone with medical qualifications in charge.") The balance of the activities in the mixed economy will be considered by MRC in an internal review early next year.

With more than ten years at Unilever, the new secretary is likely to forge closer links between MRC and industry. "We have a duty to transfer more ideas into industry, which will be easier if a better understanding with key companies develops", he says. This may generate extra funds as a by-product.

There is also a need for better links between the council's own research units and universities, says Rees. Small, isolated units neither have the interactions that are necessary for the best research nor contribute as they should to the educational environment. One advantage of the planned move of the Clinical Research Centre to the Royal Postgraduate Medical School in Hammersmith is that the MRC scientists will be physically associated with an educational institute. The same would apply to the National Institute for Medical Research in the fullness of time.

One immediate priority for MRC is the chronic shortage of graduate students in biomedical science, which will do nothing to help the existing shortage in high-quality postdocs.

In the past year, says Rees, MRC has recorded its first shortfall in the uptake of postgraduate grants. He puts this down to lack of motivation fuelled by the declining morale of the scientific community. An increase in the value of postgraduate grants may restore interest, he suggests.

Peter Newmark

French maths

Paris

HALF of France's top-level mathematicians will retire in the first decade of the next century and recruitment of young mathematicians is far too slow to provide replacements. The figures, released at a meeting of the two major French mathematics societies at Palaiseau last week, are particularly galling as France prides itself on being the third country in mathematics after the United States and the Soviet Union. France has six Fields medallists.

A boom in recruitment in the 1970s was followed by a dramatic decline in the 1980s with the result that 1,130 out of 2,280 university mathematicians are aged between 42 and 49, with only 139 under 35. Recruitment is running at 55-60 posts a year, whereas 75-100 are needed to ensure that a shortage will not develop early next century.

Peter Coles

Corporate power in universities opposed

Tucson, Arizona

A WASHINGTON-BASED organization that helped to mobilize scholars against research on the Strategic Defense Initiative (SDI) has now turned its attention to fighting what it sees as excessive corporate influence in academic life.

The National Coalition for Universities in the Public Interest was founded in 1984 with the help of consumer activist Ralph Nader because of concern that pro-consumer faculty members were being dismissed and Nader's sources "were drying up", explains David Noble, professor of history at Drexel University in Philadelphia.

During the past two years, the group has helped to organize two conferences for anti-SDI activists and now works with the Boston-based Committee for Responsible Genetics to discourage research related to biological warfare. It is sponsoring several studies on corporate 'invasion' of universities and helps faculty members who believe they have been dismissed because of their views.

Leonard Minsky, executive director of the coalition, believes corporations have become more involved in universities

because of increased concern about US economic competitiveness.

Two federal acts have made close links between universities and corporations more attractive. In 1980, the Small Business University Patenting Act took a first step by allowing universities to hand out exclusive rights for the manufacture and distribution of new products developed through publicly funded research. Previously, exclusive rights were not granted.

The Research Development Loan Program further stimulated corporate involvement by providing tax concessions for companies that financed university research. Minsky believes that basic science suffers as more emphasis is put on short-term, applied research.

The group maintains a telephone hotline, which it says receives about 20 calls a day from faculty members concerned about job security because of their political views.

By late January, the coalition expects to publish a litigation handbook that includes a list of resources by state and of where to go for help as well as an academic defence strategy for faculty in trouble.

Elizabeth Pennisi

Progress towards a national consensus for the SSC?

Denver, Colorado

WITH the process of site selection for the Superconducting Super Collider (SSC) well on its way but signs of money to build it apparently retreating, backers of the huge particle accelerator are concerned not to let political support dwindle as the number of possible sites comes down to a shortlist and then a single eventual winner. But the majority opinion at a meeting of scientists, industrialists and politicians



in Denver was not to delay final site selection until after authorization of construction, but rather to elicit unconditional expressions of commitment to the SSC as a project of national importance.

The Denver meeting, sponsored by the American Physical Society with the enthusiastic help of the state of Colorado and its governor, brought together the Secretary of State for Energy, the present as well as the former presidential science adviser, four state governors, one lieutenant governor, three Representatives and a Senator, as well as several influential scientists and science administrators. A prevailing optimistic note, sounding the importance of the SSC for US science, technology and education, was punctuated by warnings concerning the precarious state of current federal budget negotiations, and punctured by Senator Pete Domenici (New Mexico) who, appearing via satellite from Washington, said he saw little chance of Congress spending between \$1,000 and \$2,000 million a year over the next few years, as long as automatic Gramm-Rudman budget reductions were in force.

Correction

The Japanese attack on Pearl Harbor, Hawaii, was in 1941 (see *Nature* 330, 515; 1987). □

But other speakers, notably John Herrington, Secretary for Energy, and George Keyworth, former science adviser to President Reagan, restored the audience by declaring that the SSC could inject new energy into the ailing research base of the United States and thus boost industrial competitiveness.

Some partisan groups were in evidence, particularly prominent being 'Farmers for the SSC in California', who were at pains to stress that local opposition in their state was due to a vocal minority. But lawsuits

have been filed to keep the SSC out of California, and a cloud undoubtedly hangs over that bid.

Amid the cheerleading, Herrington had a confused message on international collaboration, saying that the United States should seek financial help from outside, but at another point remarking that New York's proposed site crossing into Canada was dismissed from the competition because "this is an American project". The slogan of the meeting was "can America afford not to build the SSC?", and the issue of international help, which the Department of Energy has never wholeheartedly pursued, seemed to be an unwelcome guest at the party.

David Lindley

UK nuclear operators unhappy with radiation dose advice

London

NUCLEAR power operators in Britain are understood to be still smarting from the decision of the National Radiological Protection Board (NRPB) not to delay publication of new guidelines on radiation dose limits which are likely to form the basis of more stringent safety regulations.

Last month, the NRPB published interim guidelines based on preliminary reassessment of the incidence of cancer in survivors of the atomic bomb blasts at Hiroshima and Nagasaki (see *Nature* 330, 30; 1987). Some sections of the nuclear power industry are thought to be unhappy that the NRPB has issued guidelines in advance of an expected recommendation by the International Commission on Radiological Protection (ICRP). When it became known that the NRPB was to publish interim guidelines, private, high-level representations were made to the NRPB, but were apparently ignored.

The NRPB recommends that the permitted exposure of workers be cut from 50 millisieverts (mSv) to 15 mSv, and that of the general public from 1 mSv to 0.5 mSv.

The Central Electricity Generating Board, which operates Britain's nuclear power stations, already imposes a dose limit of 10 mSv on its workforce, and so does not anticipate any problems in instituting new regulations. British Nuclear Fuels (BNFL), which operates the reprocessing plant at Sellafield, says that plant and machinery operating since the mid-1970s would be able to meet any new safety requirements. Measures are being discussed that would reduce risk to some 1,000 workers (about 10 per cent of the workforce) who may be exposed to doses greater than the suggested 15 mSv limit. The United Kingdom Atomic Energy Authority (UKAEA) regards the research upon which the NRPB bases its advice as inconclusive; radiation dosage at

UKAEA facilities is kept "as low as reasonably practicable". Of about 8,000 UKAEA radiation workers, 200 received doses in excess of 15 mSv in 1986.

The Health and Safety Commission is currently re-examining radiation control measures in the light of the NRPB's advice, and has set up a Working Group on Ionizing Radiations; any recommendations are likely to take the form of a voluntary code of practice.

Henry Gee

More nuclear safety

London

IN a mood of post-Chernobyl safety consciousness, the world's nuclear power station operators have joined forces to "improve the flow of information and maximize safety". The Association of Nuclear Operators comprises more than 150 electricity utilities from 30 countries, producing more than 99 per cent of the world's nuclear-generated electricity.

Since the association was launched in Paris in October, it has been agreed to establish regional centres in Atlanta, Moscow, Paris and Tokyo, with a small coordinating centre based in Austria, Switzerland or Britain.

The association, which will work closely with the International Atomic Energy Agency, is a forum for the exchange of information, to be achieved largely by teams of engineers visiting foreign power stations to observe operating procedures.

According to Lord Marshall, chairman of Britain's Central Electricity Generating Board and the new association's steering committee, emphasis will be placed on "man-machine interactions", where there have been "important differences of philosophy in the past between Europe, the United States and the Soviet Union".

Simon Hadlington

Controversial proposals for reform at Tokyo University

Tokyo

MOVES are afoot for radical reform of the graduate school of Tokyo University, Japan's leading national university. Under a plan put forward by the Faculty of Science, the undergraduate and graduate schools would be fused to create a six-year master's course. But other faculties and institutes connected to the university are resisting the proposal, which some critics say will only enhance the 'elitism' of the university. Behind the dispute lies a struggle for greater independence by the university's research institutes.

The science faculty's plan calls for students to have two years of general education (as at present) followed by four years of specialization leading to a master's degree, instead of the present system of a four-year undergraduate course followed by two years in graduate school. Science faculty members argue that 80 per cent of their students already go on to graduate school, and by fusing the system they will be able to offer a better course. But the faculties of law and economics oppose the plan as most of their students leave university at the undergraduate level.

Akito Arima, vice-president of the university and former dean of the faculty of science, who is one of the prime supporters of the scheme, says that the graduate school has always been treated as an appendix of the undergraduate system, and has been starved of funds and facilities. In contrast, the various science institutes, whose professors also belong to the graduate school, have been comparatively well-treated, he says. The proposed new graduate school is intended to draw in new funds, technical staff and laboratory space for researchers.

But the professors at the university's institutes are suspicious that the scheme is a plot to steal graduate students, their "golden eggs", says Arima. With a six-year course, they fear that students will tend to stay on in the faculty's laboratories rather than move to the institutes. Professor Michio Oishi of the university's Institute of Applied Microbiology is also concerned that the new scheme may cause the already limited influx of graduate students from other universities and overseas to dry up completely, enhancing the elite characteristics of the university.

Arima dismisses these fears as a "misunderstanding". He says that under his proposal, graduate student intake would be increased by 20 per cent and there would be plenty of graduate students for everyone. And graduate students from other universities and overseas would be able to enter the system at the fourth-year

level by taking an examination, just as they do now.

The proposal comes at a time when some of the National Research Institutes for Joint University Use, such as the High Energy Physics Laboratory in Tsukuba, are advocating the establishment of a 'graduate university'. These national institutes are among the best equipped in Japan, but they suffer from a lack of graduate students — the new graduate university would supply their needs. Arima says that his faculty's proposal is not intended to compete with the 'graduate university'.

What is particularly galling is that Tokyo University's research institutes are steadily gaining independence, and are one by one converting to the status of National Research Institutes for Joint

University Use. First to go was the Institute of Space and Astronautical Sciences, and it will be followed by Tokyo Astronomical Observatory and the Institute of Solid State Physics. The proposal to establish a new graduate school seems intended to save the university from break-up.

In March, professors of the science faculty unanimously adopted a resolution to review the graduate school system. Following this decision, the university as a whole decided to request a budget in fiscal year 1988 to investigate the proposal to establish a 'postgraduate-oriented' university. A decision on the budget request must be made by the Ministry of Education, Science and Culture by the end of this year. If the budget request is granted, although small, it will be a symbolic indication that some sort of reform lies ahead. But in Japan, where consensus must be reached, it is unlikely that changes will be as radical as those proposed by the faculty of science. **David Swinbanks**

Bleak prospects for Hungarian science as research is cut

London

THE Hungarian Ministry of Finance is planning major cuts in the basic research budget, in spite of recent government assurances to the contrary, according to Academician Ivan T. Berend, president of the academy. Speaking on Budapest Radio's main current events programme, *168 Hours*, Berend said that the cuts will almost certainly mean some academy institutes and university departments will have to close. He sharply attacked the "ideological concept" prevalent in the Ministry of Finance, which, he said, holds that health, education and research are "not directly productive" sectors of the economy, and represent a burden on the budget which has to be reduced.

During the past eight years, Berend said, Hungary's scientific infrastructure has deteriorated considerably because of lack of resources and of investment funds. Many of the instruments now in use, he said, are almost obsolete. Since 1985, the

Academy of Sciences has been expressing its concern, and, after prolonged negotiations, a national scientific research fund was established in 1985, which allotted 4,000 million forints to basic research for the 1986–90 five-year plan, a net increase of 10 per cent.

The Hungarian economy, however, is facing considerable problems, and in July this year the government introduced major spending cuts. It was announced, however, that in distributing such funds as were available, priority would be given to the research, health and education sectors. Only a few weeks ago, Berend said, the academy and other research institutions, including the universities, were given guarantees that their funding would not be significantly affected. Although some money would have to be "siphoned off" in 1987, it would be made up to them in the future, and, furthermore, the real value of the 1988 research budget would be maintained. But the new cuts, Berend said, will amount to a 30 per cent reduction over 1987–88, and, in view of previous government assurances, the academy and research institutes were left without contingency plans.

The Academy of Sciences has 36 research institutes, employing almost 3,000 researchers. It is also partially responsible for the financial support of 80 university departments. Cuts on the scale proposed by the Ministry of Finance will make it impossible to maintain this research network, Berend said, as across-the-board cuts would mean that everyone would be left with insufficient funds to function properly. **Vera Rich**

Einstein manuscript

Washington

ONE of Albert Einstein's oldest manuscripts on the theory of relativity has been sold to an unidentified bidder at an auction at Sotheby's in New York for \$1.2 million. The 72-page manuscript, written between 1911 and 1912, had been expected to sell for more than \$700,000 (see *Nature* 330, 97; 1987). It set a US record as the highest price paid at auction for a manuscript, and set a worldwide record for unillustrated text manuscripts. **Carol Ezzell**

Historians, whigs and progress

SIR — Edward Harrison's critique of anti-whiggery in the history of science (*Nature* 329, 213; 1987) presents a strangely paradoxical discussion of historical inquiry. He describes the anti-whig as a baconian who attempts to investigate the past "with an observant but empty mind". According to this caricature, while historians catalogue the "synchronic events of a particular period", only scientists see "ideas diachronically . . . having an effect on later periods".

There is an implicit assumption that the only ordering principle of the history of science is the progress of scientific ideas, but this is precisely the point for which Butterfield justly castigated the whig historians. Harrison approvingly quotes Butterfield's criticism of the whig habit of drawing lines between events that "converge beautifully on the present", but overlooks Butterfield's concern that the whig historian "comes to imagine that [this line] represents something like a line of causation".

The whig idea of progress transforms the growth of liberal democracy, or of modern science, from a process to be explained into an explanatory device. But Harrison's great river of science is not just a stream to be mapped at increasingly finer scale; the historian wishes to explain why it takes particular channels and why it sometimes abandons one channel to follow another. For this we must investigate not only its internal dynamics, but also those other factors, the geological structures or vegetative cover, that shaped its flow.

Thus, when anti-whigs document scientists' "family and social lives, travels, accomplishments, publications and awards", these are not isolated facts; rather they mediate and influence the transmission and development of scientific ideas and as such are the meat of historical explanation. The historians' concern with how social, political, religious and other factors shaped scientific institutions and ideas leads to a quite different kind of history from one that focuses on an autonomous cascade of scientific ideas.

Harrison is also troubled lest historians, whose training leaves them ignorant of science, write a history devoid of science and abandon the difficult modern fields for the safer realms of antiquarianism. As to the latter concern, from 1969 to 1986 the portion of items dealing with the nineteenth and twentieth centuries in the chronological section of the *Isis* Critical Bibliography of the History of Science increased from 46 per cent to 54 per cent, while all other historical periods suffered a decline. One could equally well, or equally fallaciously, argue that this re-

flects an unwillingness to contend with the serious historical problems of sources, languages and interpretation that confront the student of early science.

As to the allegation that historians have somehow abandoned the study of that "network of interrelated experiments and theories", one can only note that discussions of the content of science continue to appear, often on the same pages as discussions of its context. To appropriate a line from Harrison, "there is no reason why a history of science cannot combine both views", because as presently practised, it does.

STEPHEN C. MCCLUSKEY

*Department of History,
West Virginia University,
Morgantown, West Virginia 26506, USA*

1. Butterfield, H. *The Whig Interpretation of History*, 12 (Horton, New York, 1965).

SIR—Edward Harrison's Commentary attacking "priggish" historians of science assumes that scientists inevitably take a 'whiggish' view of how their discipline has developed. He implies that the blind alleys of scientific investigation are deservedly forgotten. Historians who deliberately try to revive interest in now-rejected theories offer a relativist view of knowledge unacceptable to the scientist. Such claims misrepresent what most historians of science are trying to do and leave scientists free to preserve their blinkered view of the past.

Most historians of science are well aware of the need to show how certain theories became incorporated into the framework of knowledge — only a small contingent of radicals would argue that there is no empirical basis for the success of some theories over others. The problem with whig history is that it often underestimates how long the scientists took to decide what the 'right' position should be. Historians look at 'wrong' (that is, subsequently rejected) theories not to eliminate the distinction between right and wrong, but because we cannot understand how the right one triumphed unless we know why some scientists were at first tempted by the wrong ones.

My own research on non-darwinian evolution theories would probably be classed as a good example of prig history, but it is not my intention to diminish the success of modern darwinism. I merely insist that we cannot ignore the development of evolutionism in the crucial years between the *Origin of Species* and the emergence of the modern synthetic theory, when most biologists did not accept natural selection. The whig history of evolutionism dismisses Lamarckism and other non-darwinian concepts as minor aberrations, but in so doing it (1) ignores

the work of a whole generation of biologists, (2) endorses an oversimplified view of the cultural impact of evolutionism and (3) creates artificial puzzlement over questions such as 'Why was Mendel ignored?' The priggish historian looks at the 'false' theories dismissed by the whig interpretation and shows that Mendel was ignored because a whole generation of biologists accepted a non-darwinian and non-mendelian view of evolution and heredity.

Scientists have a legitimate right to focus on certain lines of theoretical development which hindsight tells us were important, but all too often this technique has been used to dismiss as irrelevant everything that was not on the 'main line'. Historians of science oppose whig history because they know that if we do not take the 'blind alleys' seriously, we shall never understand the real structure of the main line itself.

PETER J. BOWLER

*Department of History and
Philosophy of Science,
The Queen's University of Belfast,
Belfast BT7 1NN, UK*

Undoing apartheid

SIR—J.G. Wilson (*Nature* 328, 288; 1987) is misinformed. As your survey (*Nature* 327, 259; 1987) makes clear, black South Africans are not "denied the possibility of a scientific education and career". I visited South Africa recently and can confirm that there are many "non-white" students at the University of the Witwatersrand. And a student who worked here this summer received his degree in electrical engineering from the University of Stellenbosch, in spite of his being classified as "coloured" by the South African government.

In my view, Professor Phillip Tobias and his like-minded colleagues, who make no "secret" of their abhorrence of apartheid, should be invited to international meetings to receive gold medals for the work they are doing to dismantle apartheid in South African Universities.

ROGER BURNSIDE

*Arecibo Observatory,
PO Box 995,
Arecibo, Puerto Rico 00613, USA*

Credit where it's due

SIR — I note with interest and wonder at the reasons why authors in your journal and others so seldom acknowledge help obtained from library and information services while according fulsome praise to typists, photographers, graphics artists, technical assistants, colleagues and practically everyone else.

JUDITH PALMER

*Rothamsted Experimental Station,
Harpenden, Herts AL5 2JQ, UK*

Melting is merely skin-thick

The phase transition from solid to liquid, perhaps the most puzzling of all, probably begins with surface-melting where the crystallographic orientation is appropriate.

As phase transitions go, melting is an obscure phenomenon, at least compared with, say, vaporization. Both liquids and solids, being always in equilibrium with their vapours, will, like any condensed phase, disappear into the gas phase given a sufficiently good surrounding vacuum (although that may take a very long time). Liquids, by contrast, appear not to co-exist below the melting-point with their solid phases, which serve as nucleation centres for crystal growth. Yet in the absence of nucleation, supercooled liquids are commonplace while superheated solids are not (whence the use by chemists of melting-point measurements as a means of characterizing pure compounds).

These distinctions are not particularly profound in that they are the consequences of spurious circumstances. Condensed phases will evaporate into a good vacuum, for example, because their molecules can get physically away. Moreover, there is ample evidence of what may be called pre-melting as solids are heated towards the melting point; the concentration of interstitial atoms, or of dislocations, can be shown experimentally to increase with increasing temperature inversely with some power of its offset from the melting point, lending support to the notion that melting is just another order-disorder phase transition.

But how is pre-melting to be visualized? As the accumulation of lattice disorders whose effect is to distribute disorder uniformly through the bulk of a crystal? Or as the appearance of possibly evanescent bubbles of liquidity within an ordered crystalline array? One obvious comment on the second possibility is that most liquids are less dense than the solid phases from which they are formed, whence the common supposition (for which the Soviet author H.J. Frenkel was apparently first responsible) that melting is, to begin with, a surface phenomenon.

Here is a neat set of measurements that puts that supposition on a more solid foundation. B. Pluis and three colleagues at the FOM-Institute for Atomic and Molecular Physics at Amsterdam, in the Netherlands, have now described in the most convincing way a set of observations supporting the supposition that the onset of melting is a surface phenomenon — by showing that surface melting is a predictable (or, at least, a *post hoc* rationalizable) event (*Phys. Rev. Lett.* **59**, 2678; 1987).

What they have done is to show that, at pre-melting temperatures, the disordering of the surface structure is sensitively a function of the crystallographic direction in the underlying lattice of a carefully machined single crystal of metallic lead.

How does one set about revealing such a phenomenon? First find a suitable probe, best visualized as a physical device (say a hat-pin), whose penetration of the supposedly solid object of the study is sensitively a function of its orientation. For practical reasons (as even the most slender hat-pins are too thick for comfort), the authors of this study have used beams of 75.6 keV protons (for practical purposes, billiard balls at that energy) which are known to be virtually unable to penetrate oriented lead crystals when fired at them in the [101] direction, which coincides with the orientation of parallel linear arrays of lead atoms. For a well-ordered lead crystal, the scattering of protons in a roughly perpendicular direction is only small — but is also highly sensitive to the blocking effect of the ordered array of lead atoms.

What the group at Amsterdam has done is to machine a tiny cylindrical sample from a single crystal of lead whose orientation is almost accurately known. They have mounted their piece of lead in such a way that it can be moved transversely to the plane defined by the blocking direction [101] and the nearly-perpendicular direction [121], the former of which is defined by the proton beam and the latter by that of the line to the proton detector.

As the geometry of the experiment is defined, different distances away from the thickest peak of the lead cylinder correspond to particular crystallographic directions, while it is possible to calculate the average penetration depth (in atomic spacings) along the [101] direction at each point. For all places on the surface of the cylindrical sample, the penetration of the proton beam appears to have increased rapidly within 100 degrees of the melting-point (roughly 600 K), from less than two atomic spacings to nearly 10. More significantly, the penetration depth seems to be exquisitely sensitive to the crystallographic orientation of different patches on the surface of the lead crystal with respect to the underlying crystallographic axes.

Nobody will be much surprised by this. The parts of a solid crystal at which the number of atoms per unit surface area are the smallest are likely to be those at which

disorder will most easily set in, or be observed. The interest in what Pluis *et al.* have done is in relating their observations of proton scattering to earlier broodings about "surface free energy", as it is called. Here again the analogy between evaporation and melting is appropriate.

It is common, in the discussion of the thermodynamics of the equilibrium with each other of small liquid droplets, to take the view that the free energy of atoms in a droplet and in bulk liquid is the same, but that the formation of droplets may entail additional "surface free-energy" (which is nothing other than the diminished aggregate capacity of a small drop to hold onto peripheral atoms because the van der Waals forces do not saturate). That explains why large drops often grow at the expense of small by transfer through the vapour phase.

If the surface is considered as a whole, this is nothing other than a way of saying that the total free energy per atom is modified by some small quantity proportional to the cube root of the square of the total bulk. Pedantically, however, none of this allows the legitimate definition of a single free energy of a water molecule in a water droplet, or of a lead atom in a small crystal, which is then modified for the small structure as a whole by a surface-free energy (per unit of area) for some hypothetical surface. Nor does it allow the definition, per atom, of the free energy of a single atom in the surface of a crystal with a particular orientation.

But it is a legitimate subject for inquiry to ask whether the total energy of the microsystem would decrease if the atom were at one place rather than the other. The unconscious error is that of mistaking equilibrium processes, of which free energy is the determinant, for dynamic processes such as those involved in evaporation and, for that matter, melting.

What Pluis *et al.* have nevertheless shown is that there is indeed a correlation between computed surface free-energy per atom of atoms in surfaces of different orientation, as derived from vapour pressure measurements of lead crystals, and the likelihood that surfaces of the same orientation will show disordered behaviour near the melting point. That also points to the need for a better understanding of the dynamics of pre-melting. Perhaps people will in the end be thrown back onto molecular dynamics.

John Maddox

Cell motility

New jobs for dynein ATPases

Ian R. Gibbons

STUDY of the molecular basis of cytoplasmic motile systems requires the use of an appropriate *in vitro* model. Although various models are available that retain different levels of the *in vivo* functional and regulatory mechanisms, one of the most useful for studying the elementary properties of a motile system has been direct observation, in the light microscope, of individual protein filaments gliding over the surface of a glass slide coated with the appropriate energy-transducing protein¹. Study of this phenomenon using purified dynein and brain microtubules is described by Bryce Paschal *et al.* on page 672 of this issue², as well as by other reports that have appeared during the past two months³⁻⁶. These data provide the long-awaited validation of cytoplasmic dyneins as cytoplasmic motors distinct from the axonemal dyneins of cilia and eukaryotic flagella⁷, and demonstrate that the motile repertoire of dynein is significantly more elaborate than believed hitherto.

The diversity of the numerous microtubule-associated proteins has become highly confusing to the non-specialist. A helpful clarification can be obtained by grouping them into classes and families (see table), analogous to the general scheme proposed for actin-binding proteins in an earlier News and Views article⁸ by Tom Pollard. Because many forms of cytoplasmic microtubules exist in a state of dynamic equilibrium⁹, the study of the skeletal (non-energy-transducing) class of microtubule-associated proteins has generally focused on their ability to co-polymerize with tubulin *in vitro* and to stabilize the resultant tubules. Little is known about their physiological function in the cell.

The members of the energy-transducing class of microtubule-associated proteins are easier to study by virtue of their ATPase activity and of the various *in vitro* motility models. An abundance of new evidence indicates that most of the complex bidirectional movements of vesicles and organelles that characterize a living cell are driven by one of two molecular motors interacting with a microtubular track. Cytoplasmic microtubules are dynamic, polarized structures, with relatively rapid polymerization of tubulin occurring at their 'plus' end and concurrent slower polymerization or depolymerization at their 'minus' end. In most systems, including nerve axons, the plus end of a microtubule corresponds to its distal end (that further removed from the cell centre) and the minus end to its proximal end, with the same polarities applying also

to the more stable tubules of flagellar axonemes. The tubule motor responsible for distally directed vesicle movement in most cells seems to be kinesin, an ATPase that was first isolated from nerve axons and shown to produce both unidirectional movement of latex beads along a microtubule and gliding of microtubules on glass¹. The motor responsible for proximally directed movement, dynein ATPase, was first characterized from

There remains some uncertainty about the premise of a separate protein motor for each direction of transport along microtubules. Most notably lacking is an explanation for how latex beads coated with a dynein-like protein from the giant amoeba *Reticulomyxa* can mimic the bidirectional motion of the *in vivo* vesicular transport on microtubules in reticulopodia of this organism¹⁴. Nevertheless, the basic premise appears likely to be valid. In flagellar axonemes, where dynein produces sliding between pairs of tubules it does so by forming a structurally stable bond to one tubule and a dynamic cross-bridging attachment to the other. In the retrograde transport of cytoplasmic vesicles, the corresponding structural

Characteristics of microtubule-associated proteins					
Class	Family	Source	Subunit M_r (K)	Structure	<i>In vitro</i> movement
Energy-transducing	Dyneins	Cilia/flagella	>400	2-3 globular ATPase domains joined by slender stems	Proximal
		Brain	>400		Proximal
	?	<i>Reticulomyxa</i>	>400	Rod with branches at one end and small fork at the other	Bidirectional
		Kinesins	110 134		Distal Distal
Skeletal	MAP 1	Brain	~350	Thin filament projecting from tubule	None
	MAP 2	Brain	~270	Thin filament projecting from tubule	None
	Tau	Brain	55-62	Globular, heterogeneous in size	None
	Tektin	Cilia/flagella	45-55	Similar to intermediate filaments	None

Microtubule-associated proteins (MAPs) are grouped into classes by their function and subdivided into families according to their physical and enzymatic properties. This grouping is arbitrary and will probably be modified as more data become available. Dynein from brain was formerly called MAP 1C. Subunit M_r gives the relative molecular mass of the principal heavy polypeptide subunit; in many cases additional smaller subunits are also present. Right-hand column: 'proximal' indicates movement of an attached structure toward the proximal (-) end of microtubule, corresponding to retrograde axonal transport; 'distal' indicates movement of an attached structure toward the distal (+) end of microtubule, corresponding to anterograde axonal transport.

ciliary axonemes as long as 20 years ago, and its structural and enzymatic properties are relatively well known¹⁰, although as mentioned above, a functionally active form of cytoplasmic dynein was reported only recently^{3,4}. The properties of the axonemal dyneins that constitute the 'outer arms' of the microtubules of cilia and sperm flagella are particularly well known¹⁰; a well-defined structure seen in the electron microscope, very large polypeptide subunits of $M_r > 400,000$ susceptible to vanadate-mediated photolysis; and high sensitivity of ATPase activity and motility to inhibition by vanadate. The present paper by Paschal *et al.*² demonstrates that the gliding microtubule procedure can be used to assay the function of outer-arm dynein from sperm flagella, and this provides a basis for probing functional domains of dynein in a manner similar to that successfully exploited with myosin^{11,12}. Not surprisingly, the direction of force generated by axonemal dyneins, which has been determined by observations on the sliding disintegration of trypsin-treated cilia, is proximal¹³, the same as that of cytoplasmic dynein from brain⁴.

attachment of the dynein presumably has to be to the membrane of the vesicle, and is possibly mediated by membrane-bound tubulin¹⁵. Outstanding questions concern the regulation of the motors and how vesicles become attached to the correct motor. Regardless of how the answers to these questions turn out, it is now evident that dynein fulfils a greater variety of jobs in the cytoplasm than has previously been realized. □

1. Vale, R. D., Reese T. S., & Sheetz, M. P. *Cell* **42**, 39-50 (1985).
2. Paschal, B. M. *et al.* *Nature* **330**, 672-674 (1987).
3. Paschal, B. M., Shpetner, H. S. & Vallee, R. D. *J. Cell Biol.* **105**, 1273-1282 (1987).
4. Paschal, B. M. & Vallee, R. B. *Nature* **330**, 181-183 (1987).
5. Vale, R. D. & Toyoshima, Y. Y. *J. Cell Biol.* **105**, 96a (1987).
6. Lye, R. J. *et al.* *Cell* **51**, 309-318 (1987).
7. Hyams, J. S. *Nature* **330**, 106 (1987).
8. Pollard, T. D. *Nature* **312**, 403 (1984).
9. Mitchison, T. & Kirschner, M. *Nature* **312**, 237-242 (1984).
10. Johnson, K. A. *A. Rev. Biophys. biophys. Chem.* **14**, 161-188 (1986).
11. Harada, Y. *et al.* *Nature* **326**, 805-808 (1987).
12. Toyoshima, Y. Y., Kron, S. J., McNally, E. M., Niebling, K. R., Toyoshima, C. & Spudich, J. M. *Nature* **328**, 536-539 (1987).
13. Fox, L. A. & Sale, W. S. *J. Cell Biol.* **105**, 1781-1787 (1987).
14. Euteneuer, U. *et al.* *J. Cell Biol.* **105**, 125a (1987).
15. Stephens, R. E. *Biol. Cell* **57**, 95-110 (1986).

Ian R. Gibbons is at the Pacific Biomedical Research Center, University of Hawaii, Honolulu, Hawaii 926822, USA.

Superconducting ceramics

Specific heat clues for theory

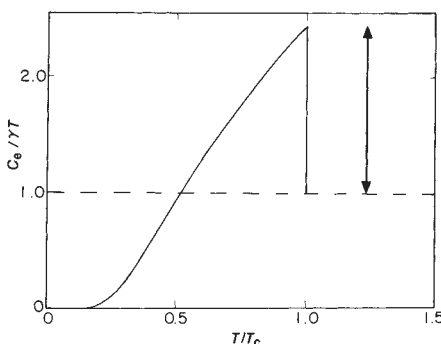
R. A. Fisher, J. E. Gordon and N. E. Phillips

SINCE the 90-K superconducting transition was first observed¹ in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ (YBCO), various phenomena suggestive of superconductivity at much higher temperatures, frequently in the range 200–240 K, have been reported. More recently, there has been a growing recognition of the importance of both the oxygen content and the ordering of the oxygen vacancies in determining the superconducting properties of the new materials. On page 637 of this issue², Lægheid *et al.* report calorimetric evidence for an ordering process that occurs near 220 K and is correlated with the occurrence of superconductivity. The magnitude of the specific-heat anomaly at 220 K in different samples is approximately proportional to the 'discontinuity' in specific heat at 90 K, the latter being taken as a measure of the completeness of the transition to the superconducting state. Although different interpretations of the anomaly at 220 K are possible, it seems clear that it will provide a clue to the nature of the superconductivity in this interesting and important system.

Measurements of specific heat have generally been regarded as the ultimate test for distinguishing bulk superconductivity from superconducting filaments, which might occupy only a small fraction of the sample volume. Measurements of resistivity or, to a lesser degree, of magnetic susceptibility can be misleading in this respect, but a measurement of the specific heat gives the average of that property over the whole volume of the sample. The electronic specific heat, C_e , of a BCS (Bardeen, Cooper and Schrieffer) superconductor that exhibits a sharp transition to the superconducting state is represented in the figure. In the normal state, realized in fields that exceed the critical field, the specific heat C_{en} depends on temperature (T) as γT , where γ is a constant proportional to the density of electronic states at the Fermi level (the approximate energy of electrons that play a significant role in the electronic properties). In the superconducting state, there is a discontinuity, $\Delta C(T_c)$, at the critical temperature, T_c , and at lower temperatures the specific heat, C_{es} , has an approximately exponential dependence on temperature. The linear term characteristic of the normal state is absent and $(C_e/T) \rightarrow 0$ as $T \rightarrow 0$. The exponential temperature dependence of C_{es} reflects the condensation of the electrons into Cooper pairs and the consequent development of a gap in the density of states.

In principle, the completeness of the

superconducting transition could be determined either by measuring $\Delta C(T_c)$ (which is proportional to the fraction of the sample that goes superconducting) or by looking for a residual linear term $\gamma'T$, in zero applied field (which is proportional to the fraction that remains normal). Both methods depend on a knowledge of γ . The former is rendered uncertain, even for BCS superconductors, because $\Delta C(T_c)$ can vary from $1.43\gamma T_c$, for the weak-coupling limit, to as much as $2.5\gamma T_c$ in cases in which the coupling is strong. (The coupling strength is the strength of the electron-phonon interaction which, in the BCS theory, determines the ratio of



The electronic specific heat of a BCS superconductor in the normal and superconducting states. Broken line, $C_{en}/\gamma T$; solid line, $C_{es}/\gamma T \approx 8.5(T_c/T) \exp(-1.44T_c/T)$ at low temperatures. Arrow shows depth of discontinuity at $T=T_c$, $\Delta C(T_c)/\gamma T_c = 1.43$. This value can increase to 2.5 in the strong coupling case.

T_c .) For the new high- T_c materials in particular, there are several problems associated with the application of these criteria.

First, in addition to C_e , the quantity measured always includes the lattice specific heat, C_l . Because T_c is so high, C_l makes by far the larger contribution at T_c and $\Delta C(T_c)$ is only of the order of 1 per cent of the total specific heat. This difficulty in determining $\Delta C(T_c)$ can be alleviated by measuring C both in zero magnetic field and in a field large enough to suppress the superconducting transition by several degrees. Second, the transitions observed to date are relatively broad, and $\Delta C(T_c)$ must be determined by a construction similar to that in Fig. 2 of the paper² by Lægheid *et al.*, on page 637. Third, at low temperatures the critical field is very high, of the order of 100 T, and γ has not been measured directly.

Furthermore, for the new superconducting oxides there are particular reasons why a linear term in the specific heat in zero field is not necessarily an indication of an incomplete transition to the superconducting state. The electron-

phonon coupling, generally believed to be the basis for the attractive interaction between electrons that leads to superconductivity in all other materials ('heavy-fermion' compounds possibly excepted), may not be able to account for the superconductivity of the new materials.

Other mechanisms are being considered. For one of these, the resonant-valence-bond (RVB) model³, the superconductivity does not involve a gap in the density of electronic states, so that the temperature dependence of C_{es} differs from that of the BCS theory, a linear term in C_{es} being an intrinsic property of the superconducting state. Another possible origin of a linear term is the existence of 'two-level systems' (TLS) as suggested by sound-velocity data^{4,5}. These are associated with double minima (two stable positions) in the potential energy of an atom, presumably oxygen atoms in YBCO, between which tunnelling can occur, and are responsible for the linear term in the specific heat of amorphous materials.

To date, all specific-heat measurements on YBCO and on the superconducting lanthanum-alkaline-earth-copper oxides show a linear term in the specific heat in zero field. These linear terms can be accounted for on the basis of incomplete transitions to the superconducting state, but other origins such as TLS or the gapless superconductivity of the RVB model are certainly not excluded. In the case of the lanthanum oxides, the wide variation in the values of γ' suggests that they arise at least in part from residual normal material, but for YBCO there are reports of γ' values in the range 6–8 mJ mol⁻¹ K². The small variation in γ' , apparently independent of oxygen content, might be taken as support for the RVB model. It is clear, however, that the answer to the question of the origin of the zero-field linear term must await further measurements on well-characterized samples.

Lægheid *et al.* do not report² low-temperature data and γ' is therefore unknown for their samples. The data for sample UK1 near 90 K show a transition width similar to those observed in careful measurements in other laboratories, but a relatively low value of $\Delta C(T_c)$. Their sample B39/2, for which $\Delta C(T_c)$ is substantially greater (but for which the data are not shown) seems to be more like the 'best' samples that have been studied elsewhere. It is therefore reasonable, for the purpose of comparison with the 220 K anomaly, to take their values of $\Delta C(T_c)$ (ΔC_p in their notation) as a measure of the extent or quality of the superconducting transition.

Lægheid *et al.* discuss² possible origins of the 220-K specific-heat anomaly, and conclude that the most likely explanation lies in the ordering of the oxygen system. As they note, there is now considerable evidence that it is ordering of the oxygen

in the Cu–O chains that is important for superconductivity. Furthermore, it is clear that high oxygen content favours good superconducting properties. The authors do not propose a model that relates the ordering to their specific-heat results, but as they observe a larger 220-K anomaly for the better samples, one must infer that the entropy change associated with ordering is greater in the samples with the higher oxygen content. But because the samples with lower oxygen content have more oxygen vacancies, one might naively have expected a greater entropy change for those samples.

The antiferromagnetic ordering in other copper oxides affords another interesting comparison with the 220-K anomaly. It is known^{6,7} that CuO orders magnetically at about 230 K. (Because the molal entropy change associated with the ordering in CuO is about the same as that associated with the 220-K anomaly in YBCO, it is not possible to attribute the anomaly to CuO impurities.) Furthermore, there is evidence for antiferromagnetic ordering in La₂CuO₄ at similar temperatures^{8,9}, although the Néel temperature (the critical temperature for antiferromagnetism) is a strong function of oxygen content. For both La₂CuO₄ and CuO, the magnetic moment per copper atom is about 0.4–0.5 Bohr magnetons, much less than the moments associated with conventional antiferromagnetic ordering in ionic copper salts. Despite the differences in chemical composition and crystal structure among these compounds, and despite the lack of clear-cut magnetic evidence in the case of YBCO, the possibility of antiferromagnetic ordering as the origin of the 220-K anomaly in that material should be considered.

Further investigation is necessary to confirm whether it is oxygen ordering, antiferromagnetism or some other ordering phenomenon which is responsible for the anomaly observed by Lægsgreid *et al.* More investigation will be necessary before we can be certain. There is little question, however, that the discovery of the 220-K specific-heat anomaly in YBCO provides another piece of evidence which will be useful in unravelling the riddle of high-temperature superconductivity. □

1. Wu, M.K. *et al.* *Phys. Rev. Lett.* **58**, 908–911 (1987).
2. Lægsgreid, T., Fosshelm, K., Sandvold, E. & Julsrud, S. *Nature* **330**, 637–638 (1987).
3. Anderson, P.W. *Science* **235**, 1196–1198 (1987).
4. Golding, B. *et al.* *Phys. Rev. B* **36**, 5606–5608 (1987).
5. Collocott, S.J. *et al.* *Phys. Rev. B* **36**, 5684–5686 (1987).
6. Hu, J.-H. & Johnston, H.L. *J. Am. Chem. Soc.* **75**, 2471–2473 (1953).
7. O'Keefe, M. & Stone, F.S. *J. Phys. Chem. Solids* **23**, 261–266 (1962).
8. Vaknin, D. *et al.* *Phys. Rev. Lett.* **58**, 2802–2805 (1987).
9. Mitsuda, S. *et al.* *Phys. Rev. B* **36**, 822–825 (1987).

R.A. Fisher, J.E. Gordon and N.E. Phillips are in the Materials and Chemical Sciences Division, Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, USA. Permanent address for J.E.G.: Physics Department, Amherst College, Massachusetts 01002, USA.

Tumour necrosis factor

Another chapter in the long history of endotoxin

Lloyd J. Old

LIKE all interesting molecules, tumour necrosis factor (TNF) can be viewed in many ways. First, its anti-cancer properties are being investigated in many centres. Second, TNF is linked to various toxic manifestations of infectious, neoplastic or autoimmune diseases. For instance, it has been implicated in the profound weight loss (cachexia) associated with chronic parasitic or bacterial diseases as well as with cancer¹, the circulatory collapse and shock associated with acute bacterial infection^{2,3}, the death of animals with cerebral malaria⁴, and the development of acute graft-versus-host disease⁵. The study of Kevin J. Tracey *et al.* on page 662 of this issue⁶ provides further evidence for the key role of TNF in septic shock associated with gram-negative bacteraemia.

A third reason for interest in TNF is the growing understanding of the role it plays, with other polypeptide mediators, in initiating and regulating inflammatory and immunological responses, and in controlling the growth of haematopoietic and other cell types. Twenty or so such polypeptides have now been identified, and in most cases cloned, and members of the family include interferons, TNF and lymphotoxin, interleukins (IL-1 to IL-6), colony-stimulating factors, and a series of growth factors such as TGF- α and TGF- β . The molecular language created by these factors is turning out to be very complex. Factors like TNF and IL-1 have an astonishing range of activities, with an influence on the growth, differentiation, function or movement of virtually every cell type tested⁷. There is considerable overlap in the activities of the factors, with two or more structurally unrelated molecules eliciting the same biological endpoint. Many of the factors were first thought to be the product of a single cell type, but it is now becoming the rule that factor production is not cell-type specific and that the same factor can be produced by cells of multiple distinct lineages. Mediator interaction adds to the complexity, with one factor eliciting the production of itself or another factor, and factors potentiating or antagonizing the actions of each other.

Attracting attention

These three aspects of TNF — anti-tumour action, role in inflammation and immunity, and toxicity — are precisely the activities that attracted so much attention to bacterial endotoxin, the pyrogenic and

toxic component of the cell wall of gram-negative bacteria. In fact, the main issues confronting workers in the TNF field are familiar to devotees of endotoxin. Endotoxin is a lipopolysaccharide (LPS), and it is the lipid A component of LPS that is responsible for its broad range of biological activities⁸. Although macrophages release TNF in response to several different agents, LPS is the most potent TNF-releasing factor known. The fact that TNF and LPS have such similar actions has led to the view that TNF is one of the chief mediators of LPS action.

LPS, like TNF, has beneficial as well as harmful effects. For instance, LPS protects mice against bacterial infections and lethal doses of X-rays, and TNF can do the same. TNF also mimics the therapeutic effects of LPS on experimental tumours — tumour haemorrhagic necrosis and tumour regression.

On the other hand, TNF in higher doses can induce a shock syndrome in rodents and dogs which resembles in considerable detail the toxicity of LPS^{2,9}. Pathogen-free mice are considerably more resistant to TNF toxicity than mice with chronic infections. This probably results from synergistic interactions between TNF and LPS, with low doses of LPS greatly enhancing the lethality of non-toxic doses of TNF¹⁰.

Endotoxin has long been suspected to be involved in septic shock, and administration of antibodies against the lipid A core structure of LPS has protective effects in patients with gram-negative bacteraemia¹¹. With respect to the role of TNF in septic shock, Waage *et al.*³ detected TNF in the blood of patients with meningococcal septicaemia, and found that patients with highest TNF levels were most likely to die. However, such correlations are not sufficient to assign a causative role to TNF in septic shock, as patients with parasitic diseases can have high levels of TNF in their blood¹² and cancer patients treated with TNF achieve even higher TNF blood levels without evidence of shock.

The experiment of Tracey *et al.* described in this issue⁶ provides a direct approach to determine the involvement of TNF in shock associated with bacterial sepsis. The authors found that passive administration of TNF-neutralizing antibody 2 hours before infection protects baboons against the acute lethality of 10¹¹ live *Escherichia coli*. Antibody given 1 hour before bacterial infection gives some

At its simplest, conformal equivalence can be regarded as a device of computational rather than geometrical significance; for instance, the Riemann curvature tensor of a class of conformally equivalent metrics can be written as the curvature of the parent space-time

modified by some simple terms depending on the conformal factor F . Geometries equivalent to Minkowski space are said to be conformally flat, because the parent space-time has no curvature and any non-zero components of the Riemann tensor arise entirely from the properties of the conformal factor.

But imagination and ingenuity can invest the conformal degree of freedom with more significance than this. Quantities derived from the conformal factor most naturally belong to the geometrical side of Einstein's equation, but by putting them on the other side they can be imagined to be part of the stress-energy tensor instead. This rather arbitrary procedure allows one to think of the conformal factor as a physical field, which may then be coupled dynamically to other, more familiar components of the stress-energy tensor. This may seem to the novice to be a spurious piece of business, but in its defence it must be said that it is no more spurious than the particle physicists' habit of inventing lagrangians for reasons mostly of mathematical nicety; indeed, the interpretation of the conformal factor as a dynamical quantity has the advantage that any resulting physical models are bound to have interesting geometrical properties.

After this lengthy preamble, we come to the article by Gunzig, G         and Prigogine in this issue. Gunzig and others have for some time been promoting the conformal degree of freedom as an element in the construction of models of the creation of the Universe and of quantum gravity (see Brout, R., Englert, F. & Gunzig, E. *Ann. Phys.* **115**, 78; 1978 and Englert, F., Brout, R., Truffin, C. & Windley, P. *Phys. Lett.* **B57**, 73; 1975).

In simple terms, one begins with a 'pre-cosmological' state of a Minkowski vacuum, throughout which a dynamical conformal field has a constant value. But in any such field there are quantum fluctuations, which in this case represent variable and small-scale departures of the geometry from exact Minkowski form.

The trick that Brout *et al.* have proposed in the past is that the conformal fields are also coupled to more conventional physical variables, such as the fields associated with particles similar to gauge bosons in high-energy physics. What happens then is that a fluctuation in the conformal field generates not only a region of distorted geometry but also a local density of particles. Furthermore, it can be arranged that such a fluctuation becomes self-sustaining; that is, the induced presence of matter leads to a local expansion of the space-time (for just the same reasons a conventional universe expands), which in turn causes the conformal field to grow from its original equilibrium, which generates more matter, and so on.

The net effect is that an entire expanding universe is produced from a local fluctuation in a conformal field. Conformal flatness guarantees that the universe is of Robertson-Walker form, and geometrical matching conditions dictate that it is open (negatively curved) but embedded in a flat background. In this respect the model bears a more than passing resemblance to cosmologies of the 'new inflationary' sort (Linde, A.D. *Phys. Lett.* **B108**, 389; 1982 and Albrecht, A. & Steinhardt, P.J. *Phys. Rev. Lett.* **48**, 1220; 1982). The geometrical properties of such quantum universes are in fact quite general, and similar results have been found in various models (see for example Gliner, E.B. & Dymnikova, I.G. *Sov. Astr. Lett.* **1**, 93; 1975 and Gott, J.R. III *Nature* **295**, 304; 1982).

In their contribution in this issue, Gunzig, G         and Prigogine add the further sophistication that the particles initially created by the fluctuation are black holes of about 50 Planck masses. As is well known by now, such black holes evaporate until they diminish to 1 Planck mass (about 10^{-8} kg), at which point their fate is uncertain; probably the remaining mass turns into a burst of

radiation, leaving no remnant. In this cosmology, an initial fluctuation in the conformal field leads to an exponentially expanding 'inflationary' phase, containing a small admixture of black holes. For a simple choice of parameters, the black holes evaporate and disappear at the same time that the period of rapid expansion ends, and what is left is a hot, radiation-dominated universe which evolves in the conventional way.

The particularly remarkable feature of the universe of Gunzig *et al.* is that the course of its evolution depends only on the mass of the initial population of black holes and on some fundamental constants of physics. (There was a similar outcome in Gott's model mentioned above.) It thus seems to explain the observed properties of the Universe without recourse to special pleading or carefully tuned physics. On the other hand, as Gunzig *et al.* admit, their model is simple and has yet to be joined with the familiar 'low-energy' world encapsulated in the successful standard model of particle physics. But in this it is no worse than new inflation or superstring theory. □

David Lindley is an assistant editor of Nature.

Synaptic plasticity

The role of NMDA receptors in learning and memory

Graham Collingridge

THE recognition of the involvement of the *N*-methyl-D-aspartate (NMDA) receptor in synaptic plasticity in the adult and developing nervous system has led to insights into the molecular basis of learning and memory, some of which were discussed at a recent meeting*. The widespread involvement of the NMDA receptor in synaptic plasticity is illustrated by the report on page 649 of this issue¹, where Artola and Singer show that the selective NMDA antagonist D-2-amino-5-phosphonovalerate (APV) prevents the induction of long-term potentiation (LTP) in slices of rat visual cortex. This important observation shows that mechanisms basically similar to those operating in the hippocampus² may underlie forms of synaptic plasticity in other regions of the brain (L. Bindman, University College, London).

LTP, an activity-dependent enhancement of synaptic transmission that is most pronounced in the hippocampus, is widely studied by those interested in learning and memory (see ref. 3). In this article, I shall consider separately the mechanisms

responsible for the induction of LTP, for which there is considerable agreement, from those that maintain the enhanced synaptic transmission, which are controversial.

LTP induction

In both hippocampus³ and cortex¹, it is believed that NMDA receptors participate in synaptic transmission when an appropriate pattern of synaptic inputs can transiently alleviate the voltage-dependent Mg^{2+} block of NMDA channels⁴. This may happen during high-frequency stimulation because of the prolonged depolarization provided by the temporal summation of excitatory postsynaptic potentials (e.p.s.p.) and the elevation of extracellular K^+ . Synaptic inhibition opposes this effect by shunting the excitatory synaptic current and, perhaps more important, by hyperpolarizing the cell to a level where the Mg^{2+} block of NMDA channels is enhanced. To induce LTP in visual cortex, Artola and Singer¹ needed to reduce this inhibitory influence with a low dose of bicuculline. In the hippocampus, antagonists of the neurotransmitter GABA facilitate⁵ but are not essential for the induction of LTP, perhaps because of

* Activity-dependent changes in synaptic efficiency in the vertebrate nervous system and the associated meeting of the Physiological Society, Mill Hill, London, 5-7 November 1987.

the extremely high density of NMDA receptors in this region and the lability of inhibitory postsynaptic potentials.

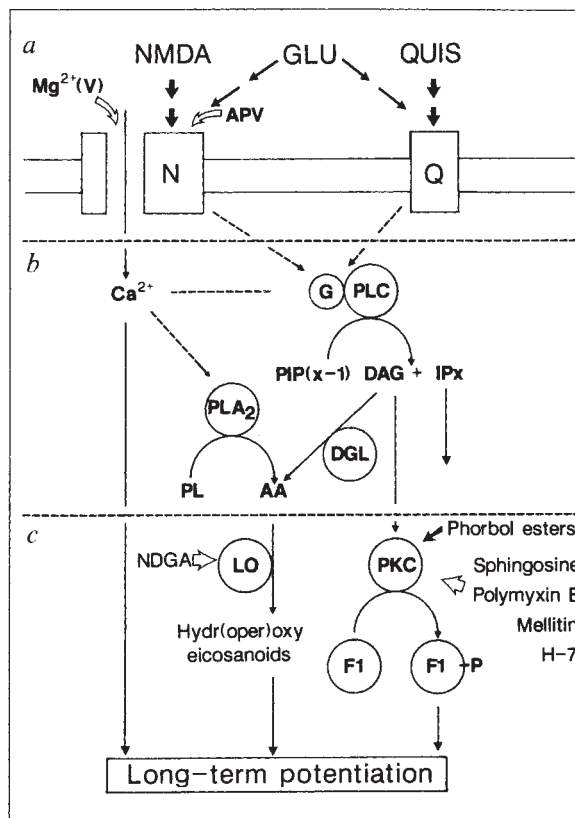
The activation characteristics of LTP, conferred by the NMDA receptor system, can be considered as 'hebbian' (T. Brown, City of Hope, Dewart, California). Thus, presynaptic activity is required to provide an NMDA-receptor agonist (such as L-glutamate), and postsynaptic depolarization, but not spiking, is necessary to alleviate the Mg^{2+} block of the NMDA channel. These gating properties of NMDA channels can account for the properties of specificity, cooperativity and associativity which make LTP an attractive model for information storage in the brain³. The requirement for both pre- and postsynaptic activity may be solely for the purpose of activating the NMDA-receptor system, as NMDA applied together with just a few low-frequency shocks², or even alone (R. Nicoll, University of California, San Francisco), can induce an LTP-like effect. (Note that NMDA receptors might be localized both presynaptically and postsynaptically in the hippocampus^{2,6}.)

LTP maintenance

There is disagreement as to whether LTP is (T. Bliss, Mill Hill, London) or is not (Y. Ben-Ari, Hôpital de Port Royal, Paris) associated with an increase in glutamate release. Evidence for the former suggestion is more fully documented, the increase being critically dependent on NMDA-receptor activation⁶. In support of a presynaptic component, LTP does not seem to be associated with an initial change in the postsynaptic sensitivity to quisqualate, an agonist believed to be selective for the receptor mediating the e.p.s.p. both before and after the induction of LTP (my own work). There may, however, be a delayed increase in sensitivity to quisqualate, which raises the possibility of a slowly developing postsynaptic component of LTP maintenance.

Calcium entry through the transiently unblocked NMDA channels, rather than through voltage-gated channels, is believed to be important for associative forms of LTP (Brown). What happens thereafter is less certain.

Several lines of evidence indicate the involvement of the phosphoinositide (PI) system in LTP. Thus, phorbol esters can either induce an LTP-like effect (Ø. Hvalby, University of Oslo) or facilitate the induction of tetanically induced LTP (A. Routtenberg, Northwestern University); inhibitors of protein kinase C prevent the persistence, but not the induction, of LTP (Routtenberg; R. Malinow, Yale School of Medicine); and LTP is associated with both a translocation of protein kinase C from the soluble to the membrane-bound form (Routtenberg



Some possible events in the generation of LTP. *a*, L-glutamate activates non-NMDA (quisqualate)-type receptors during low frequency transmission. During high-frequency transmission, the Mg^{2+} block of NMDA channels is reduced and Ca^{2+} enters the cell via these channels. NMDA may induce, and APV blocks, LTP by binding to the NMDA receptor. (Monovalent ion fluxes not shown.) *b*, Phospholipid metabolism may be somehow coupled to N- or Q-receptor activation. *c*, Tertiary events that may be involved in LTP and the possible sites of action of some drugs that promote (phorbol esters) or block these processes. N, NMDA; Q, quisqualate; GLU, L-glutamate; G, G-protein; PLC, phospholipase C; PIP(x-1), phosphoinositides; DAG, diacylglycerol; IPx, inositol phosphates; PLA₂, phospholipase A₂; DGL, diacylglycerol lipase; PL, phospholipids; AA, arachidonic acid; LO, lipoxygenase(s); PKC, protein kinase C.

and a G-protein-linked increase in PI turnover (Bliss). Protein kinase C phosphorylates a protein termed F1 (Routtenberg), an effect which is directly correlated with the generation of LTP and dependent on NMDA-receptor activation. F1 is thought to function in neural growth and development and could therefore be involved in structural changes, such as alterations in dendritic-spine morphology (P. Andersen, University of Oslo), that accompany LTP.

The metabolism of arachidonic acid by lipoxygenases yields hydroxy- and hydroperoxy-eicosatetraenoic acids, compounds that have been implicated as second messengers in *Aplysia*⁷. These eicosanoids could function as retrograde messengers in hippocampal LTP to communicate the induction, occurring postsynaptically, with a sustained increase in transmitter release⁷. It is therefore of interest that nordihydroguaiaretic acid (NDGA), an agent that inhibits certain enzymes including the lipoxygenases, blocks LTP and the associated increase in glutamate release (Bliss). The figure summarizes these receptor-operated and biochemical changes.

Developmental plasticity

NMDA receptors could be critically involved in developmental plasticity. Thus, intracortical infusion of APV renders kitten striate cortex resistant to the effects of monocular deprivation during the 'critical' period (W. Singer, Max Planck Institute for Brain Research, Frankfurt). This may involve disruption of

synaptic mechanisms during a period of consolidation, as indicated by studies with ketamine⁸, a non-competitive NMDA antagonist. It is therefore of interest that APV blocks the response of cortical neurons to visual stimuli during the critical period of young kittens more effectively than in the adult cat⁹. It could be that NMDA receptors are more prevalent, and as a consequence LTP is more easily demonstrated, in visual cortical slices prepared at this stage of development. NMDA receptors may also be involved in the fine-tuning of neural maps in optic tecta during development¹⁰.

In summary, the NMDA-receptor system has the necessary characteristics to enable it to determine the development, and further modification of the efficiency, of synaptic connections in the vertebrate central nervous system. Many more examples of NMDA receptor-mediated plasticity should soon be found. □

1. Artola, A. & Singer, W. *Nature* **330**, 649-652 (1987).
2. Collingridge, G. L., Kehl, S. J. & McLennan, H. *J. Physiol., Lond.* **334**, 33-46 (1983).
3. Collingridge, G. L. & Bliss, T.V.P. *Trends Neurosci.* **10**, 288-293 (1987).
4. Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462-465 (1984).
5. Wigström, H. & Gustafsson, B. *Acta Physiol., Scand.* **125**, 519-522 (1985).
6. Errington, M.L., Lynch, M.A. & Bliss, T.V.P. *Neuroscience* **20**, 279-284 (1987).
7. Piomelli, D. *et al.* *Nature* **328**, 38-43 (1987).
8. Rauschecker, J.P. & Hahn, S. *Nature* **326**, 183-185 (1987).
9. Tsumoto, T., Hagihara, K., Sato, H. & Hata, Y. *Nature* **327**, 513-514 (1987).
10. Cline, H.T., Debski, E.A. & Constantine-Paton, M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4342-4345 (1987).

Graham Collingridge is at the Department of Pharmacology, University of Bristol, Bristol BS8 1TD, UK.

Opioid physiology

Biosynthesis of morphine in the animal kingdom

H.W. Kosterlitz

Two years ago, I wrote a News and Views article¹ entitled "Has morphine a physiological function in the animal kingdom?". New data, reported by Charles Weitz, Kym Faull and Avram Goldstein in their paper on page 674 of this issue², provide evidence that may answer this question.

When I wrote my previous article, it was generally considered that the endogenous opioids in animal tissues are fragments of the large peptide precursors pro-opioid, pro-enkephalin and prodynorphin, whereas the non-peptide morphine is found in plants such as the poppy (*Papaver somniferum*). But about three years ago, Spector and colleagues showed³ that immunoreactive morphine is present in low concentrations in toad skin (3 pmol g⁻¹ tissue); they found other non-peptide opioids in lower concentrations in the skin of rat and rabbit and, particularly, in bovine adrenal gland and brain (0.05–0.13 pmol g⁻¹). Shortly afterwards, Goldstein and his group identified⁴ immunoreactive morphine in bovine hypothalamus and adrenal gland. They stressed that it was not clear whether the very small amount of morphine they had found was of endogenous or exogenous origin. Therefore, demonstration of biosynthesis of morphine would be decisive. This consideration is of importance because, in 1981, Hazum *et al.* found⁵ immunoreactive morphine in cow's milk and suggested a possible dietary source in animal fodder.

Last year, Spector's and Goldstein's groups both presented the results of their further analyses of endogenous non-peptide opioids. Goldstein and his colleagues found⁶ that morphine and codeine are consistently present in extracts of bovine hypothalamus but variable in extracts of bovine adrenal gland and rat brain. Even then they still could not decide whether the action of these two morphinans is of an endogenous or exogenous nature or whether the morphinans are involved in brain function.

A few months earlier, Spector and

colleagues had found⁷ free codeine and free morphine in extracts of rat brain. Because in the morphine biosynthesis of *P. somniferum*, norlaudanosoline, reticuline, salutaridine, thebaine and codeine are precursors of morphine⁸, Spector and colleagues decided to test whether intravenous injection of these morphinans would increase the free morphine content in the brain and other tissues of the rat (Table 1). It was not known whether the concentration of free morphine obtained by intravenous injection of codeine (16

vitro. This critical step of conversion is not present in rat brain or bovine adrenal gland. The authors point out that this finding is the first demonstration of the ability of an animal tissue to synthesize morphinans. This result strongly supports the hypothesis that the codeine and morphine found in mammalian tissue are of endogenous origin.

Although the concentrations in animal tissues of free endogenous codeine and morphine are much lower than the concentrations of opioid peptides (Table 2), some speculation on their significance may be permissible. There are two significant differences between the peptide and non-peptide compounds^{10–12}. First, morphine interacts almost exclusively with the μ -opioid receptor, having a relative affinity of 0.98 (maximum = 1.0; Table 2) but [Met]enkephalyl-Arg-Arg-Val-NH₂ is less selective, having a relative affinity of 0.66

Table 2 Peptide and non-peptide opioid compounds with preference for the μ -site

Compounds	Relative selectivity for μ -binding in guinea-pig-brain ¹¹	Content (pmol g ⁻¹)	Tissue
[Met]enkephalyl-Arg-Arg-Val-NH ₂	0.66	6.5	bovine hypothalamus ¹²
[Met]enkephalyl-Arg-Phe	0.60	112	bovine hypothalamus ¹²
Morphine	0.98	0.03 (free) 0.7 (total)	rat brain ^{7,9}

Selectivity index: K_i^{-1} for $\mu/(K_i^{-1}$ for $\mu + K_i^{-1}$ for $\delta + K_i^{-1}$ for κ).

pmol g⁻¹; Table 1) is high enough to exert an agonist effect on the μ -opioid receptor. Spector and his colleagues have now shown⁹ that free morphine and codeine are only small fractions of the sum of free and sulphate-conjugated compounds. In the rat brain, free morphine is 0.03 pmol g⁻¹, whereas the mean of total morphine is 0.7 pmol g⁻¹ and that of total codeine is 1.2 pmol g⁻¹.

These data indicate that it is essential to investigate whether there are in animal tissues mechanisms which convert the non-morphinan reticuline by intramolecular coupling to give the morphinan salutaridine, as occurs in *P. somniferum*. A full analysis of the fact that this conversion can take place in animal tissue is reported in the paper by Goldstein's group in this issue². These authors produce convincing evidence that in rat liver reticuline can be converted to the morphinan salutaridine both *in vivo* and *in*

at the μ -receptor and of 0.31 at the κ -receptor, and [Met]enkephalyl-Arg-Phe is also less selective, having a relative affinity of 0.60 at the μ -receptor and of 0.36 at the δ -receptor. Second, the enzymatic stabilities of non-peptide and peptide opioids interacting with the μ -receptor are also very different. Most of the peptide opioids are sensitive to the degrading action of peptidases, the exception being β -endorphin; in contrast, morphine is much more stable. We still know very little about the physiological role of endogenous morphine and codeine and their correlation with the functions of opioid peptides. □

Table 1 Morphine concentration (pmol g⁻¹) in some rat tissues after intravenous injection of saline or three morphinans (from ref. 7)

Tissue	Saline	Salutaridine (31 nmol g ⁻¹)	Thebaine (10 nmol g ⁻¹)	Codeine (17 nmol g ⁻¹)
Brain	0.026	0.059	0.096	16
Liver	0.011	0.136	0.150	115
Intestine	0.017	0.099	0.112	182

Conversion to free morphine is much less with salutaridine or thebaine than with codeine.

1. Kosterlitz, H.W. *Nature* **317**, 671 (1985).
2. Weitz, C.J., Faull, K.F. & Goldstein, A. *Nature* **330**, 674–677 (1987).
3. Oka, K., Kantrowitz, J.D. & Spector, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1852 (1985).
4. Goldstein, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5203 (1985).
5. Hazum, E. *et al. Science* **213**, 1010 (1981).
6. Weitz, C.J., Lowney, L.I., Faull, K.F., Feistner, G. & Goldstein, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9784 (1986).
7. Donnerer, J., Oka, K., Brossi, A., Rice, K.C. & Spector, S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4566 (1986).
8. Kirby, G.W. *Science* **155**, 170 (1967).
9. Donnerer, J. *et al. J. pharmac. exp. Ther.* **242**, 583 (1987).
10. Goldstein, A. & James, I.F. *Molec. Pharmac.* **25**, 343 (1984).
11. Kosterlitz, H.W. *Proc. R. Soc. B225*, 27 (1985).
12. Sonders, M. & Weber, E. *J. Neurochem.* **49**, 671 (1987).

H.W. Kosterlitz is Director of the Unit for Research on Addictive Drugs, University of Aberdeen, Marischal College, Aberdeen AB9 1AS, UK.

Molecular diagnostics

Genetics cracks bone disease

Bryan Sykes

It is rare these days to find an inherited disease where the transition between the old and new genetics has been comparatively painless. The familial pattern is a molecular biology *Blitzkrieg* that brushes aside cherished ideas, mechanisms and even personalities in its path. The collagen diseases have escaped this particular trauma because, thanks to luck as much as anything, the old ideas have turned out to be right. By and large it is the collagen *aficionados* themselves who have led the attack that has elevated brittle bone disease (osteogenesis imperfecta or OI) from an arcane orthopaedic riddle into a highly developed example of mutational cause and effect, as was revealed at a recent meeting*.

OI, which affects about 5,000 people in the United Kingdom and a proportionately similar number worldwide, can be a devastating disease. At worst, bone strength is so reduced that victims die at birth from the fracture and collapse of the rib cage when the first breath is drawn. At the other end of an almost continuous spectrum the disease is so mild that the tendency to fracture is barely noticeable. In between these two extremes it can be a savagely crippling condition.

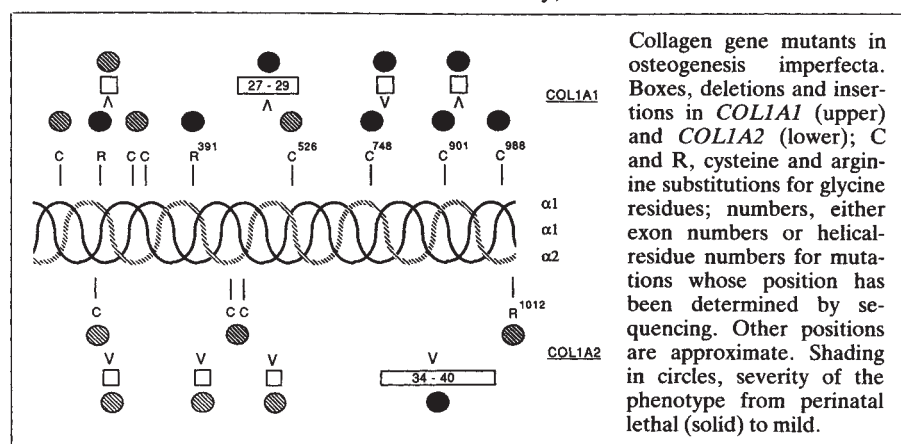
Collagen, weight for weight as strong as steel and used as the reinforcing rods in bone construction, turns out to be encoded by a gene that is remarkably feeble, collapsing under the slightest mutational pressure. And, as was discussed at the meeting, these mutations cause OI. Although there are many molecules claiming to belong to the collagen family (a purely administrative definition — collagens are molecules discovered in collagen laboratories) it is the common-or-garden variety, collagen 1, that does the work in bone. This molecule is a triple helix with the three polypeptides so tightly wound that only glycine is small enough to occupy the axial position. To achieve this, the polypeptides are 338-fold tandem repeats of the tripeptide Gly-X-Y, residues at the X and Y positions often being proline and hydroxyproline. Each collagen 1 molecule contains two identical chains (α_1), and a third (α_2) which is very similar (see figure). The α -chains are encoded at the unlinked loci *COL1A1* and *COL1A2*, respectively, but the general organization of the two genes is almost identical. The α_2 helix is encoded by 42 exons, each of which is a simple recombinational descendent of a 54-base pair (bp) ancestor. A single fusion of two 54-bp

exons in the α_1 gene is the only difference between the two. These two highly conserved features, glycine positioning and multiple exons, offer many opportunities for damaging mutation.

One of the eight possible results from substituting a single base in a GGC glycine codon is a cysteine residue, normally absent from the helical region. When this occurs in *COL1A1*, a proportion of the molecules becomes disulphide bonded and the reducible dimers are easily seen on polyacrylamide gels. This result, first seen 2 years ago in a patient with a perinatal lethal form of the disease, initiated a search for similar mutants, and a dozen successes were described at the

substitution in the molecule: the closer it is to the carboxy terminus, the more severe it is. At positions 988 (B. Steinmann, University of Zurich), 901 and 748 (P.H. Byers, University of Washington, Seattle) in the 1,014 residue α_1 -chain, the phenotype is lethal. If the substitution is at position 526, the patient survives but with severe bone deformity. Arginine seems to be an even more damaging substitute than cysteine, with a lethal phenotype resulting from a mutation at position 391 (J. Bateman, University of Melbourne).

Mutants in the α_2 -chain are far less severe than their counterparts in α_1 because, for reasons of stoichiometry, only 50 per cent of molecules incorporate the abnormal chain rather than 75 per cent, as is the case with an α_1 mutant. A Gly-Arg substitution in the last triplet in the α_2 -chain, at position 1,012 (R.J. Wenstrup, University of Washington, Seattle), produces a moderate bone deformity, whereas the same mutation in α_1



meeting. The substitutions affect helix formation and stability (see the recent News and Views article by Robert Freedman, *Nature* 329, 196; 1987). The helix winds from the carboxy terminus and progresses until it meets the mutated glycine. It takes some time to come to terms with the interruption during which time the unfolded portion of the chains continues to be modified by prolyl and lysyl hydroxylases. Eventually, folding restarts and carries on to the amino terminus, leaving behind an unwound section surrounding the Gly-Cys substitution site. These molecules have reduced thermal stability, but it is unclear whether it is this or other quality-control mechanisms that lead to their destruction. In any event, the mutant molecules have difficulty leaving the rough endoplasmic reticulum and fibroblasts are dilated and engorged. This cellular constipation also affects other secreted molecules such as fibronectin and collagen 3. A proportion of the mutants do escape and probably go on to wreak further organizational havoc in the highly ordered collagen fibril.

There appears to be a gradient of effect which depends on the position of the

would almost certainly have been lethal.

Deletions of one or more exons are also turning up quite frequently. Two multi-exon deletions, one from *COL1A1* (27-29) (D.J. Prockop, Thomas Jefferson University, Philadelphia) and one in *COL1A2* (35-41) (M.C. Willing, University of Washington, Seattle) are lethal but the effect of smaller deletions, involving usually only a single exon, depends, like the glycine substitutions, on which chain is involved and where. Again, mutations towards the 3'-end of the helical domain produce the most severe effects and deletions from *COL1A1* are more serious than from the equivalent position in *COL1A2*. Because exons encode integral triplet 'cassettes' they can be deleted without interrupting the helix folding, as the glycine residues amino-terminal to the deletion are still in register. What is disturbed, though, is the distribution of charged residues in the X and Y positions which is normally precisely matched between α_1 and α_2 chains. Another effect is that the unequal lengths produce a ragged amino terminus to the helix, disturbing the recognition site of the protease that cleaves the globular propeptide from

* 3rd International Conference on Osteogenesis Imperfecta, Pavia, Italy, 1-3 October 1987.

that end of the precursor. The mechanisms for generating deletions are genomic recombination and mutations at intron-exon boundaries. Flush excision of exon 19 in one case and of exon 28 in another are the surprising results of mutations affecting the 5' splice junctions — in simpler systems these mutations would probably have produced a messenger RNA with intron transcripts and the attendant frameshifts and/or premature terminations. Does this hint at a different mechanism for editing polyexon genes?

The sheer size of the collagen genes has dissuaded investigators from blanket sequencing to determine specific mutations, and all the examples discussed here started life as observations of abnormal protein behaviour, which in turn concentrated efforts at sequencing on small regions of the genes. Thus, the defined mutations are a strategically biased sample and many predicted mutations, such as Gly-Val substitutions would remain undetected. Most thoughtful groups are now working hard on their own ingenious solutions to more efficient mutant detection.

If the detailed characterization of mutants has by necessity been slow, biased and sporadic, the contribution of restriction-fragment length polymorphism segregation analysis to OI has been far more general. All centres with access to large dominant OI pedigrees are collaborating to enlarge the number of families linked to *COL1A1* and *COL1A2*. Only 2 of 42 families discussed are not completely concordant with one or other of the loci (P. Tsipouras, University of Connecticut; my own work). If these recombinants are confirmed it means that, although unlinked families exist, their frequency is low enough for accurate prenatal diagnosis on small families based on identification of the concordant collagen 1 allele in a chorionic villus sample.

There is no doubt that the enormous amount of information about the protein and its genes amassed in the past 15 years has helped the interpretation of the recent genetic data and that without it the inherent problems of a two-locus disorder would have been very hard to fathom. This continuity is not only the strength but also the weakness of OI research. Comfortable in a world where phrases like 'heterogeneity' and 'complex interactions' were excuses for lack of progress, some of the old guard showed an unnerving enthusiasm for the rare exceptions to the unifying simplicity which was the meeting's message. The bewildering range of OI phenotypes are all mutants of the two collagen structural genes, and although incomplete in many important respects, the relationship between cause and effect is at last beginning to make sense. □

Bryan Sykes is in the Nuffield Department of Pathology, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

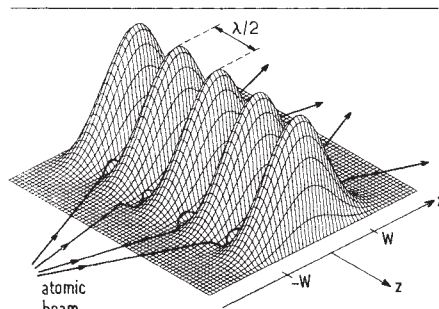
Optics

Laser manipulation of atoms

A. Ashkin

SINCE the first demonstration of the trapping of neutral atoms in a focused beam of light¹ (optical trapping) last year, there have been many developments in the study of the interactions of light and matter. Some of these were discussed recently in a News and Views article by Juha Javanainen². Other developments in techniques for trapping and manipulating atoms, dielectric particles and even organisms using light forces were reported at a recent conference*.

The subject of optical trapping started in 1970, when I demonstrated the stable optical trapping of dielectric spheres by radiation pressure³. It was then realized that similar resonant forces should be strong enough to affect significantly the



The potential energy of atoms in a beam encountering a transverse standing laser beam. The laser beam is detuned to the blue, so the atoms are repelled from the antinodes and drawn to the nodes of the standing wave, causing channelling. (From ref. 7.)

thermal motion of atoms⁴. Over the next ten years an understanding of the basic forces on atoms developed. Scattering forces arise when photons are scattered (resonantly absorbed and re-emitted) by atoms or particles, delivering momentum in the process. For each photon scattered by a sodium atom, the atomic speed changes by 3 cm s^{-1} . Gradient forces are caused by perturbations of atomic energy levels by strong applied fields, pulling atoms into, or pushing them away from, regions of high intensity.

Experiments showed that atomic beams can be stopped by scattering forces arising from thousands of photons from a counter-propagating laser. As the atoms slow down, their energy levels have to be shifted by a spatially varying magnetic field, or the laser frequency continuously retuned (chirped), if they are to continue absorbing the photons.

Atoms in a slowed beam can then be cooled in 'optical molasses': pairs of counterpropagating lasers directed along

each axis are tuned to longer wavelengths than the atomic transition ('to the red'). As atoms move in any direction, their Doppler shift causes them to feel an opposing (damping) force from the scattering forces. In the experiment last year, we showed that sodium atoms cooled to $250 \mu\text{K}$ (the quantum limit resulting from the intrinsic width of the optical transition) could then be trapped by gradient forces within a single focused laser beam¹. We also showed in a similar trap that micrometre- and submicrometre-sized dielectric particles could be trapped and levitated.

Two groups have since cooperated to build a large-volume, magnetically modified scattering-force trap called 'magnetic molasses' (E. Raab, Massachusetts Institute of Technology (MIT); M. Prentiss, AT&T Bell Labs). They use weak, spatially varying Zeeman splitting of the atomic energy levels to avoid the restrictions of the optical Earnshaw theorem, according to which scattering-force traps cannot work if the force is proportional to the laser intensity. More than 10^7 atoms were confined in a volume 0.5 mm across (to give densities of about $10^{11} \text{ atoms cm}^{-3}$) at temperatures of about $600 \mu\text{K}$ for more than 2 min. Atoms leak from this trap at a rate which, at high densities, is predominantly proportional to the square of the number of trapped atoms (Prentiss). This implies that the cross-section for two-body collisions between trapped atoms is large at low relative velocity. The non-linear losses limit the densities that can be achieved in these traps, which could frustrate those looking for collective phenomena such as Bose condensation.

An optical trap of moderate volume using two lasers has been constructed at the National Bureau of Standards (NBS) (P. Lett). This molasses-cooled trap relies on the combined action of scattering and gradient forces. It has been used to measure the associative ionization of two excited sodium atoms ($\text{Na}(3p) + \text{Na}(3p) \rightarrow \text{Na}_2^+ + e^-$) at low temperature. The cross-section at 1 mK (giving a mean velocity of only 100 cm s^{-1}) is about $1 \times 10^{-13} \text{ cm}^2$, 1,000 times larger than at 300 K . Calculations⁵ of the cross-sections for dimer formation predict interesting quantum resonances in the molecular formation rate.

A single-frequency diode laser has been used to cool caesium atoms in optical molasses (C. Weiman, University of Colorado). The use of diode lasers, frequency-locked by feedback from an external Fabry-Perot resonator, represents a major simplification in technique.

* Optical Society of America, Rochester, 18–23 October 1987.

Densities of about 5×10^7 atoms cm^{-3} at a temperature of 125 μK (the quantum limit for caesium) have been achieved with virtually stopped beams and multiple loading of the trap. This is the coldest vapour yet produced and has an average velocity of about 15 cm s^{-1} . Weiman also presented preliminary evidence of a new optical-pumping scattering-force trap using weak magnetic fields.

Several experiments that use atomic beams slowed by the scattering force are in progress. A Zacharias fountain, in which slow atoms rise and fall under gravity through a resonant radio cavity, has been constructed using a beam of sodium atoms cooled by the chirping technique (M. Zhu, Joint Institute for Astrophysics). The fountain will be used for high-resolution spectroscopy of hyperfine levels in the ground state, observed using 'Ramsey' fringes. The width of these fringes is reduced as the effective cavity residence time increases, so that the fountain (with a residence time of about 100 ms) should be very useful. Optical two-photon transitions, also with narrow intrinsic line widths, can also be studied.

A single pass of a fast (100 keV) neon ion beam through a light beam resonant with the Ne^+ ($3S-3P$) transition is sufficient to cause observable slowing (E. Riis, University of Aarhus). Calculations show that an ion beam circulating in a storage ring should crystallize as a result of the strong cooling (narrowing of the velocity distribution) possible with multiple passes: the ions, all travelling at almost the same speed, should form a stable lattice (also seen in static suspensions of charged particles) as a result of their electrostatic repulsion.

Neutral neon atoms cannot be cooled in the normal way using resonance radiation from the ground state, because the necessary short-wavelength lasers do not exist. A Japanese group has circumvented this difficulty by promoting the neon atoms into an excited metastable state and using a resonant transition from that (F. Shimizu, Tokyo University). This is the first example of cooling that does not use ground-state atoms, extending the possible applications of the technique.

An optically slowed atomic beam of sodium atoms was continuously loaded into a high-magnetic-field superconducting magnetic trap in an experiment using various techniques (V. Bagnato, MIT). About 10^9 atoms were held for minutes in a volume of 100 cm^3 and cooled by one-dimensional optical molasses to about 1 mK. The trap is suitable for studying proposals on new laser-radio-frequency cooling techniques.

A new cooling technique called 'stimulated molasses' has been used to give transverse cooling of a caesium atomic beam⁶. This effect appears in high-intensity standing waves for blue detunings,

as opposed to the red detuning of conventional low-intensity molasses. It is caused by a change in sign of the velocity-dependence of the gradient force. This effect does not saturate at high power and presumably can be used to give rapid slowing of atomic beams. The same group has also observed channelling or one-dimensional confinement of the caesium beam between the nodes or anti-nodes of an intense standing wave field⁷ (see figure).

J. Dziedzic, Y. Yamane and I have trapped and manipulated living organisms and biological cells with various shapes, using a single-beam gradient trap based on an infrared neodymium YAG laser.

We observed *Escherichia coli* bacteria and yeast cells reproducing within the laser trap and were able to separate and orientate cells, and also to manipulate organelles within the interior of living cells, and our results will be reported in next week's issue⁸. □

1. Chu, S. *et al.* *Phys. Rev. Lett.* **57**, 314-317 (1986).
2. Javanainen, J. *Nature* **329**, 16-17 (1987).
3. Ashkin, A. *Phys. Rev. Lett.* **24**, 156-159 (1970).
4. Ashkin, A. *Phys. Rev. Lett.* **25**, 1321-1324 (1970).
5. Thorsheim, H.R. *et al.* *Phys. Rev. Lett.* **58**, 2420-2423 (1987).
6. Aspect, A. *et al.* *Phys. Rev. Lett.* **57**, 1688-1691 (1986).
7. Salomon, C. *et al.* *Phys. Rev. Lett.* **59**, 1659-1662 (1987).
8. Ashkin, A., Dziedzic, J.M. & Yamane, Y. *Nature* (in the press).

A. Ashkin is in the Laser Science Research Department at AT&T Bell Laboratories, Holmdel, New Jersey 07733, USA.

Neutrino physics

Making the most of SN1987A

Terry P. Walker

THE neutrinos detected from the recent supernova 1987A (SN1987A) in the Large Magellanic Cloud by two large water Cerenkov detectors run by collaborations at Kamiokande¹ and at IMB² confirm dramatically our understanding of the generic mechanism of formation of type II supernovae: the gravitational collapse of a massive stellar core to form a neutron star or black hole. The gravitational binding energy generated is a few times 10^{53} ergs and is most easily liberated by the emission of neutrinos³. Maximum likelihood analyses of the neutrino events (11 in Kamiokande and 8 in IMB) show that SN1987A was indeed the result of such a collapse, and that the observed neutrinos were trapped on timescales of a few seconds as they radiated with a characteristic temperature of 3-6 MeV (refs 4,5).

This much, although pleasing for theoreticians, is not all that there is to discover from SN1987A. Despite the small number of neutrinos detected, their behaviour is being used to calculate many aspects of neutrino physics. How many distinct flavours of neutrinos are there? What is their mass? Can the distinct flavours oscillate (interconvert) or decay?

That neutrinos were seen at all allows the conclusion that other types of light particles were not emitted by SN1987A as copiously⁶. The duration of the neutrino signal (about 10 s) implies that a neutron star, rather than a black hole, was the final stage of the collapse (A. Burrows, personal communication), and thus the maximum binding energy which could have been released was about 6×10^{53} erg. The observed neutrinos represent a total energy of $(4 \pm 2) \times 10^{53}$ erg, and therefore the total energy carried by any flux of extra particles cannot be much more than about 3×10^{53} erg. From this argument follows an upper limit to the number of

light neutrino species^{6,7} $N_\nu \leq 6-7$. Three species are already known, associated with the electron, the muon and the tau.

Information about other particles, such as axions and other neutrino types, can also be derived from SN1987A using the facts that they would remove energy more efficiently than standard neutrinos, and that they were not observed. Any light particle that couples more weakly to matter than do standard light neutrinos will rapidly cool the supernova core and inhibit the emission of neutrinos. It is possible, however, that light particles couple to matter more strongly than standard neutrinos. These particles would be trapped during the collapse (just like standard neutrinos) and be emitted thermally with luminosities less than those of standard neutrinos. Such particles thus evade constraints from SN1987A, and one must be sure that the particles under consideration couple to matter more weakly than do standard neutrinos, which, for the cases discussed below, is guaranteed by pre-existing laboratory or cosmological bounds.

The detected neutrinos, like all known neutrinos, are all left-handed; they spin anticlockwise about their direction of motion. If right-handed neutrinos exist and are produced in the supernova as are left-handed ones, then the 'right-handed' Fermi constant must be 10,000 times smaller than the actual Fermi constant of standard weak interactions (G. Raffelt and D. Seckel, personal communication). Axions, the pseudoscalar (spin-zero, anti-symmetric) particles postulated to explain the natural presence of the CP-symmetry (particle-antiparticle symmetry) in strong interactions, could also have been emitted by SN1987A (ref. 6). They couple to nuclear matter, either directly to its constituent quarks or indirectly via coupling

100 years ago

ON THE METEORIC IRON WHICH FELL
NEAR CABIN CREEK, JOHNSON
COUNTY, ARKANSAS, MARCH 27, 1885

The Johnson County meteoric iron, the tenth whose fall has been observed, is of more than ordinary interest, because its fall is so well substantiated, because it is the second largest mass ever seen to fall, and, again, because it fell within five months of the date of the ninth recorded fall, that of the Mazapil. It is almost an exact counterpart of the Hraschina (Agram, Croatia) iron, the first of the recorded falls. On



the upper side of the meteorite ten nodules of troilite are exposed, measuring from 33mm in diameter, to 55mm long, and 25mm wide. On the lower side there are twelve such nodules exposed, 13mm in diameter, while the largest measures 19mm by 39mm. On the upper side these nodules are coated in spots with a black crust, similar to that found on the mass, but on the lower side the crust extends completely around the side of the nodules, showing the fusion very plainly. The troilite is very bright and fresh, and on the upper side one of the nodules shows deep striation, suggesting that the entire nodule is one crystal, and the exposed part is only one side of it.

From *Nature* 37, 159; 15 December 1887.

with its binding pions π^0 , with a strength that is inversely proportional to their decay constant f_A . SN1987A constrains this to be greater than about 10^{10} – 10^{12} GeV (G. Raffelt and D. Seckel, M. Turner, and K. Olive, personal communications).

The observation of neutrinos from SN 1987A also shows that the neutrinos that were emitted did not decay or oscillate on their way to the Earth, nor were they dispersed in time by more than about 1 s. Neglecting the effects of 'flavour mixing' with large mixing angles⁸, the time of flight from SN1987A to the Earth means that the life time of electron-neutrinos (of mass m) is greater than about 5×10^5 (m/eV) s. The absence of detectable γ -rays from SN1987A means that no neutrinos with masses in the range 1–100 MeV were emitted that decay radiatively with lifetimes of a few thousand seconds⁹. If neutrinos had an electric charge bigger than about 10^{-16} times that of the electron, dispersion caused by intergalactic or galactic magnetic fields would have spread their arrival times to more than a few seconds¹⁰.

The neutrinos emitted from SN1987A could not have interacted appreciably with any cosmic background of particles left over from the big bang. In particular it has been shown¹¹ that the dimensionless couplings (g) of neutrinos to massless vector (spin-one) particles must be less than 10^{-3} , to massive bosons (mass M) must be weaker than $g/M = 12 \text{ MeV}^{-1}$, and for Majoron–electron coupling must be smaller than 10^{-3} . (For comparison, g is

$1/45$ for the strong force, and $1/137$ for the electromagnetic force.) On their way out of SN 1987A, left-handed neutrinos would have been flipped to unobservable right-handed neutrinos if their Dirac masses were larger than 20 keV (G. Raffelt and D. Seckel, personal communication). It might also be possible to constrain the magnetic moment of the neutrino by demanding that helicity flipping to right-handed neutrinos near SN1987A did not occur. The effect of the interplay between this precession and matter interactions is still being debated (D. Seckel and S. Nussinov, personal communications).

With such remarkable agreement between simple theory and experiment, it is not surprising that astrophysicists and physicists alike have been seduced into trying to make the most of the sparse neutrino data from SN1987A. Their goal has been to pull out detailed features of both supernova and neutrino physics from the 19 neutrino events. Such attempts are to be viewed with caution for two reasons.

First, the observation of neutrinos from SN1987A, few though they may be, implies that type II supernovae are dominated by neutrino emission (the light emitted and kinetic energy of the shock represent only a few per cent of the total energy released and gravitational radiation even less). Thus, the neutrino flux from SN1987A should be rather insensitive to small changes in the parameters that characterize collapse and shock dynamics. Even a supernova that generates a few hundred times more events than SN1987A could be poor for diagnosing the variations in tens-of-millisecond structure expected in various detailed models of gravitational collapse.

Second, there are dangers associated with the statistics of so few events. This is exemplified by the various calculations of the neutrino mass. To do this, many authors (over 40 papers have been written) have used the relationship between a neutrino's mass m , the time at which it was emitted, t_{em} , and observed, t_{obs} , its energy E , and D the distance travelled:

$$t_{\text{em}} = t_{\text{obs}} - \frac{D}{2c} \left(\frac{m}{E} \right)^2$$

For SN1987A, D is 55 ± 15 kpc (ref. 18), c is the speed of light and I have omitted an irrelevant constant. The results of such calculations can be divided into two categories. First, exact masses for one, two or three neutrino types can be obtained by comparing t_{obs} and E of all events. Values range from a few to a few tens of electronvolts. Second, upper limits to the neutrino mass are given by examining the calculated spread of emission times and requiring it not to exceed that expected from a supernova.

Which of these values are we to believe? The first type of analysis suffers from the

sparsity of data. Apparently harmless approximations, neglecting the spread of energies for example, lead to erroneous values. The second approach is, by definition, more reliable, except that none of the calculations, I would argue, uses the correct method for sparse data, and thus cannot assess the model dependence of their mass limits. The correct method of statistical analysis for such circumstances is the method of maximum likelihood.

The likelihood function is the product of the probabilities, calculated from some model and allowing for instrumental effects, of each detected event. The emission times are calculated from the observed arrival times using the equation above, and the probabilities are calculated for each possible mass. In general, the mass limits obtained are somewhat model-dependent, but as the sparse data cannot distinguish the detailed models, any model that describes the generic behaviour of supernovae should do. Such analyses have given upper limits for the neutrino mass (at the 95 per cent confidence level) of 10–15 eV (refs 15, 16). A similar value is obtained using the Kamiokande data alone.

Are there any drawbacks to the method? Certain events could be given undue weight. The equation above shows that the lowest energy events give the best constraints. They also carry the most weight in the sample, because of their relatively low detection cross section. Because this also means that they are most likely to be background they can be removed from the Kamiokande sample, raising the mass limit to 28 eV. Until their exact status is known (the probability that they are real is 0.75), this difficulty will be unresolved.

Thus it can be seen that physics derived from the energy budget of SN1987A is reliable. Physics derived from the statistics of the 19 events is less certain but definitely useful. It is certain that both types of limits will influence the neutrino-physics community as it prepares for the next supernova. $\square$

- Hirata, K. *et al.* *Phys. Rev. Lett.* **58**, 1490–1493 (1987).
- Biota, R.M. *et al.* *Phys. Rev. Lett.* **58**, 1494–1496 (1987).
- Colgate, S.A. & White, R.H. *Astrophys. J.* **143**, 626 (1966).
- De Rújula, A. *Phys. Lett.* **193B**, 514–524 (1987).
- Bahcall, J.N., Piran, T., Press, W.H. & Spergel, D.N. *Nature* **327** 682–685 (1987).
- Ellis, J. & Olive, K.O. *Phys. Lett.* **193B**, 525–530 (1987).
- Schramm, D.N. *Proc. XIIth Recontre de Moriond* (1987).
- Miller, D.J. *Nature* **329**, 394–395 (1987).
- Friedmann, J.A., Haber, H.E. & Freese, K. SLAC Pub. 4261 (1987).
- Dar, A. & Dado, S. *Phys. Rev. Lett.* (in the press).
- Barbiellini, G. & Cocconi, G. *Nature* **329**, 21 (1987).
- Kolb, E. & Turner, M. *Phys. Rev. D* (in the press).
- Zatsepin, G.I. *JETP Lett.* **8**, 205 (1968).
- Kolb, E., Stebbins, A.J. & Turner, M.S. *Phys. Rev. D* **35**, 3590 (1987).
- Spergel, D.N. & Bahcall, J.N. *Phys. Lett. B* (in the press).
- Abbot, L.F., de Rújula, A. & Walker, T.P. *Nucl. Phys. B* (in the press).

Terry P. Walker is in the Department of Physics, Boston University, Boston, Massachusetts 02215, USA.

New superconductors in perspective

Laura Garwin and Philip Campbell

High-temperature superconductors hold much promise but are troublesome beasts.

By 2001, Japanese engineers will have developed a magnetically levitated train using the new high-temperature superconductors. Perhaps not long after, superconducting wiring will have ushered in the age of the one-wafer computer; superconducting generators and transmission lines will be set to transform power generation and distribution.

These images are those of S. Tanaka, professor of applied physics at the University of Tokyo and an adviser to the Japanese government. Professor Tanaka was speaking at the recent *Nature* conference* on "Perspectives in the New Superconductivity" and, not to be accused of thinking small, he even shared with the audience his vision of a space shuttle launched into orbit using a superconducting magnetic propulsion system.

Despite such bullishness the meeting was dominated by a sense of caution, even frustration. As speakers from around the world described recent developments, it became clear that the sheer volume and pace of research continues unabated. But much work remains to be done simply to characterize these materials reliably, sensitive as they are to the manner of their preparation and to their environment.

Interestingly, speakers agreed that the first successful applications will come, not in the areas of transportation, power distribution or digital computers that have received so much publicity, but in the field of small-scale analog devices and sensors such as the superconducting (SQUID) magnetometer.

So what are the obstacles to the revolution? According to P. Chaudhari (IBM Thomas J. Watson Research Center), whereas the range of superconductor applications would "explode" with the availability of a room-temperature superconductor, it is not at all obvious that new applications follow simply from achieving a transition temperature (T_c) higher than the boiling point of nitrogen. On the other hand, liquid-nitrogen cryogenic systems are smaller, cheaper and more reliable than their liquid-helium counterparts, so there are certainly economic gains to be made by substituting high-temperature superconductors (HTS) for conventional ones, assuming that in all other respects the new materials perform as well or better than their predecessors. Unfortun-

ately, in a few important respects they do not, as a result of what are becoming recognized as the very special properties of these copper-oxide compounds.

The distinction between extrinsic materials properties (such as the presence of grain boundaries) and intrinsic properties determined by the atomic or electronic structure of the superconductor (such as



V.L. Ginzburg, drawing an analogy with the search for a cure for cancer, recommended that all newly discovered semiconductors and metals be tested for superconductivity.

the very short coherence length) is important not only for applications, but also for understanding the nature of the superconductivity in these compounds. Both can affect the superconductivity, but only the materials properties can at least in principle be altered by appropriate processing techniques. Similarly, as M. Strongin (Brookhaven National Laboratory) emphasized, both can affect the results of experimental measurements. It is only after the contributions of each have been identified that constraints on the mechanism of the superconductivity can be derived.

Computers

The interplay of intrinsic and extrinsic properties was well illustrated in Chaudhari's discussion of computer applications. He stressed that IBM's project to develop a computer using Josephson junctions as the active circuit elements was abandoned not because of the difficulty or expense of operating at liquid-helium temperature, but because it proved difficult to fabricate the Josephson junctions reliably enough for mass production. Furthermore, Josephson junctions have lower gain than transistors, which makes them less attractive as active elements. Where the new superconductors can contribute, however, is in the conducting strips or 'interconnects' between devices or chips. These limit the speed of circuits by introducing delays which increase with the resistance and capacitance of the conductor.

Although computer speeds are not limited by these delays today, Chaudhari said, they will be tomorrow, and using superconducting interconnects will also allow increased flexibility of circuit design, as there will be no need to minimize their length.

In some respects, the new superconductors are well suited to this application. It turns out that 77 K is an ideal operating temperature for semiconducting devices: at much lower temperatures, the charge carriers introduced by doping 'freeze out' and the devices do not operate, whereas at higher temperatures the increased amplitude of lattice vibrations leads to more scattering of carriers, and hence reduced speed of operation. The large value of the superconducting energy gap in the new superconductors is also an advantage, as pulses containing higher frequencies will be able to propagate without loss.

What, then, are the obstacles? One crucial requirement for this application is a sufficiently large value of the critical current density, J_c , which is the maximum current per square centimetre that can be carried without destroying the superconductivity. Perhaps surprisingly, computer applications have the most stringent requirements for J_c —two to three orders of magnitude greater than for high-field magnets, for instance. This is because, although the currents are very low, the cross-sectional area of the conductors is tiny, and decreasing all the time as packaging density increases. Typical requirements today are $J_c > 10^6$ A cm⁻² for interconnects between devices on a chip, and $J_c > 10^5$ A cm⁻² for interconnects between chips (but this latter value will increase with time). Although R. J. Cava (AT&T Bell Laboratories) reported that critical currents of 10^6 A cm⁻² at 77 K have been achieved in orientated thin films of YBa₂Cu₃O₇ (or 'YBCO'—the most widely investigated of the 90-K superconductors), Chaudhari said that sufficient values have not yet been attained in films compatible with semiconductors.

The J_c measured in these materials is far below the theoretical value for a 'clean' type II superconductor, and arises in part from complications due to the very short coherence length of only 10–15 Å (as compared with a typical value of 1,000 Å for conventional superconductors), and in part from imperfections in the crystal lattice. The coherence length, ξ , is a measure of the distance over which the superconducting state can maintain cohe-

* Boston, 9–11 November. For an introduction to conventional and higher-temperature superconductivity and the terminology used in this report, see *Nature* 330, 21–24; 1987.

rence across a normal region; it therefore sets the size scale at which interruptions in the ordered lattice structure of the superconductor can disrupt the supercurrent. In the new materials the coherence length is comparable to the lattice spacing, which means that even a grain boundary presents enough of a discontinuity to act as a 'weak link'. Furthermore, impurities tend to concentrate at grain boundaries, making the coupling even weaker. The answer does not lie simply in growing single crystals, because the presence of a tetragonal to orthorhombic structural transition at high temperature (700 °C) in the 90-K superconductors has made it impossible to grow the orthorhombic (superconducting) material without twins, and twin planes may also act as weak links.

Happily, as Chaudhari pointed out, the same short coherence length that gives rise to weak links also provides a way to enhance J_c by allowing strong 'pinning' of the vortices formed when magnetic flux penetrates the material. The cores of these vortices are in the normal state, and their motion gives rise to dissipation, reducing J_c . The small ξ means that point defects, impurities and twin boundaries can be used to pin the vortices in place; increasing J_c to the required value may thus be a matter of materials engineering.

The other main requirement for superconducting components in computers is that their manufacture must be compatible with semiconductor fabrication technology. The ability to make thin films of YBCO has been widely achieved, but to make the films superconducting requires annealing in oxygen at temperatures in excess of 500 °C, in order to achieve the required oxygen stoichiometry. If superconductors are to be integrated into chips or packages, either the other components will have to be able to withstand such high-temperature processing, or another way will have to be found to ensure the correct stoichiometry.

Large-scale applications

The main use of superconductors in large-scale applications today is in high-field (2–20 tesla) electromagnets, in systems ranging from medical magnetic resonance imagers to particle accelerators. Other high-field applications that have been investigated include magnetically levitated trains, electric power generators and energy storage systems. Although the new superconductors have upper critical fields (H_{c2}) well above those of conventional type II superconductors (hundreds of tesla as opposed to tens), it is unlikely that they will be used at higher fields. In the first place, at fields of hundreds of tesla the Lorentz forces generated in the magnet would be sufficient to tear it apart (W. A. Little (Stanford) quoted a stress of 1,730,000 p.s.i. for a magnetic field of about 200 tesla). Furthermore, as J. K.

Hulm (Westinghouse Electric Corporation) explained, the cost of a high-field electromagnet rises more quickly than the magnetic field, making it uneconomical to increase the working field to values higher than those already accessible using conventional superconductors. Thus, although there may be good scientific reasons to exploit high fields, the advantage to be gained for applications from using the new materials is at present simply that of working at liquid-nitrogen temperature.

Hulm stressed that, even if one assumes that the new materials perform as well in all respects as the Nb–Ti alloys currently used, and are as easy to fabricate, the savings resulting from operating at 77 K instead of 4 K are measured in per cent rather than orders of magnitude because, in a large-scale system, the cryogenics constitute only a small fraction of the total cost. Nevertheless, at large levels of expenditure, savings of a few per cent are worth making.

But can the new superconductors perform equally well? Once again, the immediate problem is the critical current density. Although for these applications J_c needs only to be $>10^4$ A cm $^{-2}$, the best values for polycrystalline ceramic YBCO (recently attained at AT&T Bell Laboratories) are 7×10^3 A cm $^{-2}$ at zero field, falling to about 10^3 A cm $^{-2}$ at 1 tesla. J_c is much less in the ceramic than in thin films, partly because of the increased proportion of grain boundaries, but also because of the extreme anisotropy of J_c in these materials. Measurements on orientated thin films have revealed that J_c (and H_{c2}) varies by a factor of 30–100, depending on the direction in which the measurement is made. This anisotropy reflects the underlying crystal structure of the material, and is hence intrinsic. The value of J_c quoted above for thin films was measured in the 'good' direction; in a ceramic made up of randomly orientated grains, J_c will be reduced by contributions from the less favourable directions. Hulm pointed out that the electrical industry already uses grain-orientated steel in the cores of transformers, so it is not unreasonable to think about making grain-orientated conductors, to overcome this problem.

The ceramic superconductors are also inferior to conventional metallic superconductors in their mechanical properties. Most importantly, they are extremely brittle, which poses problems for making wires and winding magnets. For some applications, such as transmission lines and microwave cavities, Hulm suggested that a thick film of ceramic, perhaps deposited by plasma spraying, could be used. Also, Cava pointed out that YBCO is a 'good' ceramic—it can be sintered to high densities—and that ceramic processing is an advanced technology. The fact that we

may never have spools of YBCO wire is, in his view, no cause for pessimism regarding large-scale applications.

Analog devices

J. Clarke (University of California, Berkeley) surveyed the prospects for using the new superconductors in a wide range of analog devices, particularly those based on the Josephson junction. (Examples include the SQUID magnetometer, microwave and far-infrared detectors, and analog-to-digital converters.) For these small-scale devices, cooling does account for a significant fraction of the total cost. On the other hand, devices operated at 77 K will be noisier than at 4 K, so there are some applications for which lower-



P.M. Grant (IBM Almaden) commented that the use of HTS in education was already a proven and wide-scale application.

temperature operation will still be desirable. Critical current requirements in these devices are not stringent.

A Josephson junction consists of two superconducting layers, joined by a 'weak link', across which superconducting electron pairs can tunnel. The type of weak link most widely used is the tunnel junction, which consists of a thin layer of insulating oxide between superconducting thin-film layers. Unfortunately, no-one has yet succeeded in fabricating a tunnel junction from the HTS, in part because the high-temperature processing required to achieve proper oxygen stoichiometry destroys the thin oxide layer. Another way to make a weak link is with a 'micro-bridge'—a narrow strip of material which, although superconducting, achieves only weak coupling because of its small cross-sectional area. The problem here is that the length of the bridge must be small compared with the coherence length, and it is not clear that one can make microbridges only a few angstroms in length. Several groups have succeeded in making weak links, however, by taking advantage of the weak coupling between grains in the ceramic materials. Nevertheless, Clarke stressed that high- T_c tunnel junctions would be required before reliable 77-K devices could become a reality.

So far, all of the published discussion of applications of HTS has centred on the replacement of conventional superconductors in already-known applications. But as Chaudhari and Clarke both

emphasized, an important challenge will be to invent new devices that capitalize on the unique properties of these materials.

Role of structure

Although several speakers stressed the similarities between the new superconductors and other members of the superconducting 'zoo', the copper-oxide superconductors have some very special chemical and structural properties which are closely linked to their superconducting behaviour. While it was perhaps not surprising to hear the chemist C. N. R. Rao (Indian Institute of Science) describe the new materials as "chemical superconductors", the widespread recognition of this fact was reflected in physicist Strongin's statement that "we're proud of doing chemistry now... we've all got our chemistry books out". Cava stressed the special nature of the copper-oxygen bonds in these materials, in which overlap of the copper *d* and oxygen *p* orbitals allows the electrons to be delocalized, giving rise to a metallic character.

Two properties important for applications, the extreme anisotropy of H_c and J_c and the dependence of superconducting properties on the oxygen stoichiometry in YBCO, can be understood by considering the structure of the new materials, as deduced mainly from neutron powder diffraction. There are two families of copper-oxide superconductors: the '2-1-4' family, which results from doping La_2CuO_4 with, for example, strontium ($\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$), with transition temperatures of 30–40 K; and the YBCO or '1-2-3' family (in which yttrium can be replaced by almost any rare-earth element), with transition temperatures of 90–100 K. Both have layered structures, with continuous sheets of copper and oxygen atoms (the Cu–O 'planes'); the normal to the planes thus defines a 'special' direction, and the observed anisotropy reflects differences in the electronic structure parallel and perpendicular to this direction. None of the previously known oxide superconductors has such a layered structure, and this has led to speculation that reduced dimensionality may be responsible for the higher transition temperatures.

In addition to the planes, the 1-2-3 compounds have linear chains of copper and oxygen atoms, running parallel to each other in layers alternating with the Cu–O planes. Whereas in the 2-1-4 structure the two crystallographic axes parallel to the planes are equivalent, the 1-2-3 chains define a preferred direction; this reduces the crystal symmetry from tetragonal to orthorhombic. It turns out that when oxygen is removed from the structure (as happens at high temperature in air), the remaining oxygen atoms are redistributed between the sites in the chains and the normally empty sites between the chains. Eventually, these



T.M. Rice, P.W. Anderson and V.J. Emery compare models.

sites become equally occupied, thus eliminating the preferred direction, and the structure becomes tetragonal. Intriguingly, J. D. Jorgensen (Argonne National Laboratory) reported that, although the orthorhombic/tetragonal transformation may occur at different oxygen contents, depending on the procedure used to remove the oxygen, the superconductors are always orthorhombic, suggesting that the chains are important (in some as-yet undefined way) for the superconductivity.

If that is indeed the case for the 1-2-3 compounds, what about the 'chainless' 2-1-4 materials? C. W. Chu (University of Houston) cited several other differences between the two families — perhaps the most notable being the presence (2-1-4) or absence (1-2-3) of an isotope effect on T_c — which lead him to conclude, controversially, that two separate explanations will be required.

Further complications have arisen from the recent discovery of superconductivity at ~60 K in YBCO with between 6.55 and 6.75 oxygens per formula unit. Some speakers saw this as evidence for two types of superconductivity in YBCO: 'chain' superconductivity at 90 K, and 'plane' superconductivity at 60 K. Cava and Jorgensen, on the other hand, agreed that the lower transition temperature might reflect a difference in the way the oxygen atoms were ordered on the crystal lattice.

Chu and W. I. F. David (Rutherford Appleton Laboratory) suggested that additional intriguing evidence for the role of the structure in the superconductivity comes from the behaviour of various lattice-related properties at the superconducting transition. Measurements of sound velocity, acoustic attenuation, Young's modulus, specific heat and orthorhombicity all reveal anomalies at T_c , suggesting that the transition to the superconducting state is accompanied by a structural transition.

How do they work?

Seen in the context of applications, the ideal high-temperature superconductor would be identical to conventional type II materials in all respects other than T_c ; any departure from this ideal is seen as an obstacle to be overcome. For theorists, on

the other hand, these departures (insofar as they reflect intrinsic properties) reveal important differences between the mechanism of the superconductivity in the copper oxides and that responsible for conventional superconductivity.

The short coherence length, for example, is also characteristic of another class of unconventional superconductors, the 'heavy-fermion' materials, and H. R. Ott (ETH, Zürich) outlined several other similarities, including the power-law temperature dependence of specific heat and transport properties at low temperature. (In conventional superconductors these properties display an exponential temperature dependence.) The large ratio of the energy gap to T_c , now reported by several groups, may also betray a departure from the standard Bardeen-Cooper-Schrieffer (BCS) mechanism.

But measurements of the gap and other parameters have produced conflicting results. "It's a hazardous business relying on experiments, because different people do the same experiments and get different results from different interpretations, and perhaps from different materials"; thus did V. J. Emery (Brookhaven National Laboratory) summarize the theorists' predicament. Another intriguing example is provided by nuclear quadrupole resonance studies of YBCO, in which two different experimental groups have ascribed particular data to copper species in the Cu–O chains and planes respectively.

Such disagreements clearly hinder attempts to decide whether the chains or planes are the more active promoters of superconductivity. W.A. Little opts for the former, in an approach also favoured by V. L. Ginzburg (Lebedev Institute) and others, known as 'excitonic' pairing. Little was at pains to dispel one potential source of confusion: 'exciton' here refers to a variety of possible electronic excitations and not to the bound hole-electron pairs in semiconductors that the word usually represents. Virtual excitons can be exchanged by electrons to yield an effective attractive force analogous to that arising by phonon exchange in conventional BCS theory.

The excitons considered by Little are similar to those that he envisaged in organic superconductivity, in which the

necessary attraction between electrons would be provided by polarization in side-chains. For example, Little proposes that "sympathetic oscillations" can be set up in the Cu–O chains in YBCO as a result of *p* orbitals transverse to the Cu–O planes.

In support of the role of chains, Little recalled that, as Chu and others had discussed, decreasing the proportion of oxygen in $\text{YBa}_2\text{Cu}_3\text{O}_x$ simultaneously disrupts the chains and the superconducting properties. But proposing the opposite view, P.W. Anderson (Princeton) shamelessly revealed his reading habits with an analogy from the world of Agatha Christie: many murders do not require as many perpetrators. Of the distinct HTS structures, only one has Cu–O chains.

In Little's as in other excitonic mechanisms, the means by which superconductivity arises are, as reviewed by D. Pines (Illinois), entirely analogous with the electron–phonon pairing mechanism of BCS. But Pines discussed a second class of models in which electrons, as he put it, "pair first and move later" in contrast to the reverse sequence that occurs in metallic superconductivity. These models include bipolarons (in which electrons become physically paired in potential wells generated by lattice distortions) and, possibly, the resonating valence bond model of Anderson. In such models the pairing occurs independently of the condensation to a macroscopic quantum state that is the essential prerequisite of superconductivity.

Providing a starting point for such theories, a number of speakers highlighted the unusual characteristics of the normal (as opposed to superconducting) state of the HTS. T. M. Rice (ETH Zürich) drew attention to the large variety of 'ground states', including seven distinct classes of metals and insulators, that are found among the transition-metal oxides. To these have now to be added four distinct classes of 'bad metals' that are also superconductors: $\text{BaPb}_{1-x}\text{Bi}_x\text{O}_3$, $\text{Li}_{1+x}\text{Ti}_{2-x}\text{O}_4$, $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ and $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$, although only the last two fall into the novel HTS category.

Characteristically, the transitions from such states occur at relatively high temperatures, and the approach used by Rice and Anderson is to seek a formalism that explains both superconducting and normal states as a novel manifestation of strongly interacting electrons — in contrast to metals, where electron interaction energies (such as lattice coupling and Coulomb repulsion energies) are small compared to kinetic energies.

Rice is one of many theorists focusing attention on the Cu–O planes and, in particular, the copper orbitals. Rice and colleagues conclude that the normal ground state of undoped La_2CuO_4 should be an insulator for which the dominant electron energies are associated with strong Coul-

omb repulsion and kinetic energy in the form of hopping. Descriptions of this kind are conventionally known as Hubbard models.

As Rice described, spin interactions between copper atoms establish a ground state in the undoped insulator with anti-correlated spins. The spin interaction mechanism, well known in antiferromagnetic materials, is 'superexchange', in which the *d*-orbital spins on two copper atoms become anticorrelated as a result of their overlap and covalent mixing with a *p* orbital of an intervening atom of oxygen containing two electrons of opposite spin.

In this description, the doping of La_2CuO_4 with strontium adds holes to (that is, removes electrons from) the half-filled orbitals, and these holes act as charge carriers. The starting point of any Hubbard model, however, must be to predict the ground state of the undoped insulator. Observations with polarized neutrons indicate that it is an antiferromagnet, but theorists have yet to decide whether their models predict such a state or shorter-range ordering, perhaps in the form of a 'liquid' state of paired atoms having valence electrons of opposite spin ('singlets').

Both Emery and Anderson, working within a Hubbard framework, discussed a notional phase diagram for HTS in which the state is determined, for a given temperature, by the variable *x* representing the degree of doping, for example by strontium in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. Such a phase diagram was reported by S. Uchida (University of Tokyo) in a review of Japanese work. Theories suggest that as *x* increases from zero, the material changes from an antiferromagnet to an undefined (possibly spin-liquid) state and thence to the superconductor. According to Emery's model, doping with strontium adds holes to the oxygen sites in the Cu–O planes; at the superconducting transition the holes of neighbouring oxygens become paired by "enhanced superexchange" involving the intervening coppers. That is, if neighbouring coppers have anticorrelated spins then Emery suggests that it is energetically preferable for oxygen holes to form a condensate of pairs.

The notion that charge and spin can be spatially separated, first established in the case of polyacetylene by J. R. Schrieffer, was strongly emphasized by Anderson. Whatever the undoped ground state may be, doping within a two-dimensional (but not three-dimensional) system to produce the "mysterious metals" — as Anderson dubs the HTS above T_c — seems to destroy the antiferromagnetic order. According to Anderson, it produces a state consisting (in quantum-mechanical language) of the "linear superposition of all possible ways of arranging singlet pair bonds between atoms". This is the 'res-

onating valence bond' (RVB) state, 'resonating' being used here in the sense traditionally associated with the Kekulé description of benzene.

As Anderson described it, the RVB state can represent two interacting classes of electronic excitations, or solitons, within each Cu–O plane. Various calculations have supported the existence of one class: "spinons" or "chargeless fermions with spin", which can be considered as electrons with their charge removed.

The second class of solitons are positively charged and compensate for the charges added by doping; conventional holes are formed which then combine with spinons to form 'holons'. Says Anderson: "Holon solitons have nearly the full kinetic energy of the RVB state . . . holon-spinon scattering leads to the characteristic resistivity proportional to temperature over a wide range that is shown by most good high- T_c samples in the normal state".

It turns out that holons retain their identity only in the Cu–O planes. Thus although they are spinless and act as bosons, they cannot form the Bose condensate that permits the three-dimensional superconductivity observed. Anderson postulates that to move between planes "the holon must absorb a spinon and become a true hole" if spin and charge are to be conserved. The hole then tunnels into an adjacent layer to become a holon and spinon once more. Below the critical temperature the holons become paired, although the cooperative mechanism by which this occurs is yet to be fully described.

Attentive readers will have noticed the sudden increase in quotations in recent paragraphs. Much of what Anderson is saying is yet to be understood by other theorists, as Ginzburg, for example, freely admitted. "But it is an important possibility that needs to be investigated", he said. "A superconductor with a new type of ground state would be very significant."

The bewildering state of experimental and theoretical uncertainties has left theorists free to pursue their ideas virtually untrammelled by observational constraints, the principal lines of attack being phonons, excitons and Hubbard models. At the recent Berkeley meeting on superconductivity a straw poll was conducted amongst theorists as to how long it will take to resolve these difficulties. For what it is worth, the general conclusion was that it should take several years, perhaps a decade, for a full 'microscopic' theory of high-temperature superconductors to be achieved. Those seeking to exploit these materials will no doubt plough on regardless. □

Philip Campbell and Laura Garwin are Physical Sciences Editor and Assistant Editor of Nature, respectively.

The origin of the clockwork-escapement

SIR—Scrutinizing accounts of the towns of Cologne, Aix-la-Chapelle, Wesel, Arnhem, Bruges, Deventer, Ghent, Leyden, Nijmegen Rotterdam and Zutphen in the fourteenth century (transcribed and edited in the nineteenth and twentieth centuries)¹⁻⁶, I was struck by the frequency with which crossbow makers doubled as repairers of mechanical clocks. Other sources⁷⁻¹⁰, which refer to Ragusa and Lille, even mention the crossbow maker as the first local maker of a clock for the town. It is easy to ascertain from these accounts, which refer to the repairs and purchases of clocks, that these were in fact mechanical clocks¹⁰.

On further reflection, it struck me that both mechanisms under the care of this craftsman had one function in common: the capacity to store mechanical energy which can be liberated at the required moment. In the crossbow, this is effected by the released nut, in the clockwork by the escapement.

In the crossbow (Fig. 1) the nut is kept in a locked position by the trigger (Fig. 2a). By pulling the trigger the nut is set free. The force of the tight bowstring causes the nut to revolve around its axis,

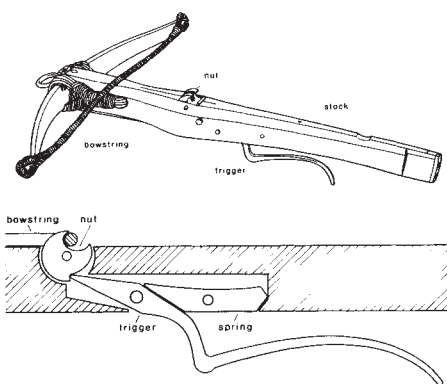


Fig. 1 Crossbow and detail of lock showing how the tight bowstring is kept in position by the nut.

leaving the bowstring free to accelerate and to push the bolt on its way (Fig. 2b).

Several versions of the clockwork-escapement exist. The best known of these is the verge-escapement, but it is not certain that this is the oldest type. The lesser-known 'semicirculus' escapement may well be older. We do know, however, that before Huygens, the oscillating element was not the swinging pendulum, but a foliot (literally, little fool) — a sort of torsion pendulum.

In the verge-escapement of the clock, the resistance of the foliot is transmitted to the crownwheel by the two pallets on the verge. The force which the crownwheel alternatively exerts on the pallets turns the verge, freeing the crownwheel to rotate under the influence of the force of the

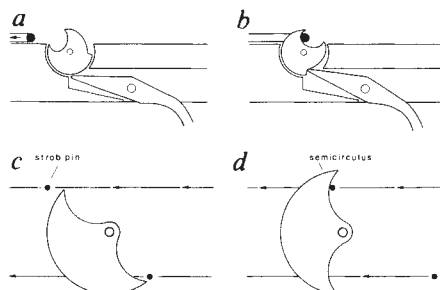


Fig. 2 Comparison of function and form of the crossbow-nut (a, b) and the semicirculus (c, d).

falling weight, which is transmitted to it via a train of gears.

In the semicirculus escapement, the resistance of the foliot is exercised by the semicirculus on the strob pins, mounted on a strob wheel. The force of the falling weight, exercised on the strob wheel via a train of wheels, makes the semicirculus revolve around its axis, freeing the strob wheel to proceed (Fig. 2c,d).

North¹¹, who was able to reconstruct this early type of escapement on the basis of the text of Richard of Wallingford¹², later discovered that it had been drawn by Leonardo da Vinci. In addition, he suggested that 'rota stroba' is derived from *stroba*, Latin for crossbow¹³. This was surmised from the resemblance of the cross-shape of verge and foliot to the shape of the crossbow¹⁴. It must be remarked, however, that the use of *stroba* for crossbow does not occur in Latin sources until the fifteenth century.

It can, in any case, be concluded that there is an obvious similarity of function between the crossbow-lock and the clockwork-escapement mechanisms. In both instances, energy is released by a revolving containment mechanism, which is moved by the force that is liberated. There is also a similarity of form of the crossbow-nut and the semicirculus, which was pointed out to me by Professor Sleeswyk. In principle, the shape of the semicirculus

is identical to that of the crossbow-nut, the only significant difference being the angle of the opening (Fig. 2).

The functional and morphological resemblance between the lock of the crossbow, with its nut, and one of the first escapement-mechanisms, with its semicirculus, is, once noticed, too obvious to be ignored. The shape of the semicirculus is as suggestive of the part the crossbow-nut may have played in the development of the clockwork-escapement as is the double employment of the crossbow maker.

I thank the late Lynn White Jr, A.W. Sleeswyk and J.M. van Winter for their suggestions and advice.

ANNIE LAUTINK-FERGUSON

B. de Beaufortweg 61,
3833 AE Lensden-Centrum,
The Netherlands

1. *De Cameraarsrekeningen van Deventer (1337-1393)* Vol. 3 (ed. van Doorninck, J.) 276, 293, 386, 414 (Deventer Boek en Steendrukkerij, Deventer, 1897).
2. *De Cameraarsrekeningen van Deventer (1337-1393)* Vol. 4 (ed. de Hullu, J.) 83 (Deventer Boek en Steendrukkerij, Deventer, 1897).
3. *De stadsrekeningen van Deventer (1394-1434)* Vol. 1 (ed. de Meyer, G.M.) 392 (Wolters-Noordhoff, Groningen, 1968).
4. *De stadsrekeningen van Arnhem* Vol. 2 (ed. Alberts, W.J.) (Wolters-Noordhoff, Groningen, 1967).
5. *Stadsrekeningen van Leiden (1390-1434)* Vol. 2 (ed. Meerkamp van Embden, J.) 24, II, 117 (Werken uitgegeven door het historisch genootschap, Müller, Amsterdam, 1914).
6. *De stadsrekeningen van Zutphen (1364-1446)* Vol. 1 (ed. Wartena, R.) 67, 122 (Gemeente Archief, Zutphen, 1977).
7. *Durđica Petrović Museji primenjene umetnosti zbornik* 13, 59-67 (1969).
8. Cipolla, Carlo M., *Clocks and Culture 1300-1700* 50, 51 (Collins, London, 1967).
9. Gaier, Claude, *L'Industrie et le Commerce des Armes dans les Anciennes Principautés belges du XIII^e à la fin du XV^e siècle* 249, 250 (Société d'Édition "Les Belles Lettres", Paris, 1973).
10. *De Cameraarsrekeningen van Deventer (1337-1393)* Vol. 3 (ed. van Doorninck, J.) 301, 336 (Deventer Boek en Steendrukkerij, Deventer, 1897).
11. North, J.D. *Richard of Wallingford* Vol. 2, 330-331; Vol. 3 p. 64 (Clarendon, Oxford, 1976).
12. Reti, L. (ed.) *The Unknown Leonardo* 249, (McGraw-Hill, New York, 1946).
13. North, *Richard of Wallingford* Vol. 1, 475 (Clarendon, Oxford, 1976).
14. Du Cange, Dominus (ed.) *Glossarium mediae et infimae latinitatis*, Vol. 6, 391 (Didot, Paris, 1846).

Physics and Fermat's last theorem

SIR—Surprisingly, Ian Stewart's description of Percival and Vivaldi's application of the theory of ideal classes to chaotic dynamics (*Nature* 329, 670-671; 1987), is not the only link between Fermat's last theorem and a problem arising in a physical context. A second unrelated and even more immediate connection was discovered recently by D. S. Rokhsar, D. C. Wright and myself (*Phys. Rev. Lett.* 58, 2099-2101; 1987) in the theory of quasicrystals, a subject familiar to readers of *Nature*.

Kummer's great work in the 1840s, referred to by Stewart, turns out to bear directly on the problem of classifying quasicrystallographic diffraction patterns. The first step in such a classification

scheme is to re-examine the fundamental enumeration by Bravais of the possible lattices of wave vectors, in the absence of the traditional crystallographic constraint that there should be a minimum distance between points. To begin such an extension of classical crystallography, one must first solve the two-dimensional lattice problem.

Two-dimensional lattices can be viewed as the ideals of the algebraic number fields called cyclotomic integers. The number of crystallographically distinct classes of such lattices with N -fold symmetry, ($N=4,6,8,\dots$) is just the number of ideal classes in the cyclotomic integers of degree N (or degree $N/2$, if N is twice an odd number). This number is unity, for all values of N less than 46, and before the connection with algebraic number theory was realized, it was hard to resist the

conjecture that the number was 1 for all values of N . A precisely equivalent conjecture was the basis for a spurious proof of Fermat's last theorem announced by Lamé and Cauchy in 1846, and refuted by Kummer's revelation that factorization into primes is not unique in the cyclotomic integers of degree 23.

For $N=46$, Kummer's discovery means that there are three distinct lattices, two of them, crystallographers will be astonished to learn, constituting an enantiomorphic pair. After that things become even wilder. There are, for example 359,057 distinct lattices with 128-fold symmetry. The computation of numbers like this is a highly non-trivial exercise, but fortunately mathematicians have been working on the problem for 140 years (it is still not completely solved), kindly saving quasicrystallographers a lot of hard work.

It is remarkable that after almost a century and a half of pristine purity, this branch of mathematical research should have found two utterly unrelated physical applications within a single year.

N. DAVID MERMIN

Laboratory of Atomic and
Solid State Physics,
Cornell University,
Ithaca,
New York 14853, USA

The 'last ribo-organism' was no breakthrough

STR—Benner and Ellington¹ emphasize the important point that primordial ribo-organisms made almost entirely of RNA and living in the RNA world² might have evolved relatively sophisticated forms of intermediary metabolism before the advent of protein synthesis. Unfortunately, the authors arrive at this correct conclusion by an argument that reveals a profound misunderstanding of basic chemistry and biology, as well as current ideas about molecular evolution.

Like others before them^{3,4}, Benner and Ellington are struck by the ubiquitous role in present-day intermediary metabolism of nucleotide cofactors such as FAD, NAD, S-adenosyl methionine, CoA and coenzyme B₁₂. We fail to understand, however, why they believe the existence of nucleotide cofactors to be evidence for metabolic complexity in the RNA world before the advent of protein synthesis. As the side chains of amino acids are chemically limited, nucleotides are far better catalysts than proteins for many reactions, particularly those involving oxidation/reduction. Thus, evolving ribonucleoprotein (RNP) or protein enzymes might have adopted nucleotide cofactors *de novo*, regardless of whether the RNA world was metabolically simple or complex.

Benner and Ellington fall into the trap of believing that any organism as adapt-

able as a virus must be completely modern because it is the product of constant streamlining. This leads them to ridicule our 'genomic tag' model for the origin of protein synthesis^{4,5} on the grounds that it is "unlikely that RNA viruses living in modern hosts contain many non-functional vestiges of an RNA world that vanished 2,000 million years ago".

As recently discussed in a News and Views article⁶, we proposed in our model that the 3'-terminal tRNA-like structures of contemporary bacterial and plant RNA viruses may be direct descendents of 3'-terminal genomic structures in the RNA world^{4,5}. This hypothesis provides the first plausible pathway for the stepwise evolution of protein synthesis, as tRNA-like structures (and perhaps specific aminoacylation) would have evolved as part of the replication machinery before the advent of protein synthesis. We also argued that several aspects of contemporary tRNA metabolism can be quite rigorously and consistently viewed as 'molecular fossils' — reactions or pathways that were used in the RNA world and persist today. The 3'-terminal tRNA-like structures are still functional today, nor would they be preserved if they were not. And the biochemistry to which we have applied the term molecular fossil is all contemporary and essential; we believe it was equally essential several thousand million years ago.

Benner and Ellington confuse the concept of a molecular fossil^{4,5} with the vernacular notion of a living fossil. For example, the fact that all organisms synthesize DNA precursors by reducing RNA precursors (using the enzyme ribonucleoside diphosphate reductase) is commonly interpreted as a molecular fossil, indicating that the pathways for synthesis of RNA evolved before the need for DNA⁷. These pathways persist because it is too late to change them; the pathways are functional, the products are essential, and the individual reactions are subject to complex regulation. Living fossil, on the other hand, is a casual term for any organism whose survival we find surprising. Living fossils are largely quirks of fate; molecular fossils survive because they are essential.

Contrary to the suggestion that many of us who think seriously about early evolution believe that the era during which RNA served as the principal catalyst in living systems was necessarily short because proteins are more versatile catalysts than RNA, neither we, nor any of the other authors cited by Benner and Ellington, have ever dared to estimate the longevity of the RNA world. We also fail to understand why any attempt to understand the origin of protein synthesis must be construed as underestimating the abilities of catalytic RNA. Our proposal that the first aminoacyl-tRNA synthetases

were made of RNA^{4,5} is clearly an affirmation of metabolic diversity and complexity in the RNA world, and we are surprised that Benner and Ellington misinterpret this proposal as revealing a mental fixation on the catalytic power of proteins.

Finally, Benner and Ellington make the serious error of assuming that the remarkable complexity of modern protein synthesis arose essential all at once, rather than one tiny step at a time. This point of view leads them to postulate a 'breakthrough' organism — the first ribo-organism to invent protein synthesis. For all its rhetorical flair, the notion of a breakthrough organism is intellectually insidious because it confuses two very different ideas — one of them obvious, the other wrong.

The obvious idea is that all forms of life on Earth must have descended from a single common ancestor or progenote⁸; no other conclusion is consistent with the remarkable similarity of all cells at the biochemical level. The wrong idea is that this progenote (reborn as the breakthrough organism?) survived because it alone was clever enough to invent protein synthesis. We prefer the more biologically plausible view that the RNA world evolved imperceptibly into an RNP world, with many interbreeding organisms contributing to the invention of protein synthesis as we know it. Whether the particular organism that we now call the progenote survived because it was better at making protein than its fellows, or simply because it did not have the bad luck to be burnt to a crisp in a volcanic eruption or swallowed up in an antediluvian deluge, is not something we are ever likely to know.

NANCY MAIZELS

ALUN M. WEINER

Department of Molecular Biophysics
and Biochemistry,
Yale University,
333 Cedar Street, New Haven,
Connecticut 06510, USA

1. Benner, S.A. & Ellington, A.D. *Nature* **329**, 295–296 (1987).
2. Gilbert, W. *Nature* **319**, 618 (1986).
3. White, H.B. III *J. molec. Evol.* **7**, 101 (1976).
4. Weiner, A.M. & Maizels, N. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7383–7387 (1987).
5. Maizels, N. & Weiner, A.M. *Cold Spring Harb. Symp. quant. Biol.* (in the press).
6. North, G. *Nature* **328**, 18–19 (1987).
7. Watson, J.D. et al. *Molecular Biology of the Gene* 4th edn, Vol. II, Ch. 28 (Benjamin Cummings, Menlo Park, 1987).
8. Pace, N.R., Olsen, G. & Woese, C.R. *Cell* **45**, 325–326 (1986).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

Questions of heaven and host

Ron Naylor

Galileo: Heretic. By Pietro Redondi. Translated by Raymond Rosenthal. *Princeton University Press*: 1987. Pp.356. \$29.95. To be published in Britain next year by Allen Lane/Peregrine.

SOME of the decisive aspects of Galileo's trial for heresy in 1633 are unknown and will always remain so; for this reason the process and its background are destined to retain an air of mystery and intrigue. One of the main reasons for this is that two quite crucial episodes were private interviews between Galileo and representatives of the Holy Office of which little is known with certainty. An essential element of practically all attempts to explain the trial and dispel the mystery has been to reconstruct — hypothetically — what came to pass in those crucial interviews. So it is with Redondi's new thesis. The book was published in Italian in 1983, and has now appeared in English translation.

It is a common theme of modern interpretations of Galileo's trial that the available facts and documents of the official record do not account for the action taken against him. Attention is often directed to the fact that he had after all obtained the necessary *imprimature* allowing the publication of the *Dialogue on the Two Great World Systems* not once but twice. Why, it is asked, had he been treated so harshly in view of this? Suspicion has engendered belief in a plot engineered by Galileo's enemies, and in particular by the Jesuits. This in any case was the opinion of Galileo's friends in Rome at the time, and was later implied by his student and biographer Viviani.

These suspicions were strengthened by many other features of the affair, which have led to a conviction on the part of some historians that vital aspects of the trial were intentionally obscured, so that there was in effect a secret element within the trial — and it was this that proved crucial. The public and officially documented account of the trial was accordingly claimed to be the outcome of a tortuous judicial process and no more than a façade. For this reason there has long been the hope that one day some secret, lodged unknowingly within the vast records of the Vatican, might be uncovered to shed new light on the mystery. Redondi's work fits the bill precisely. We can imagine with what excitement he unearthed a 'new' document of evident

importance within those files. It revealed a wholly new aspect of the affair, the significance of which had never been recognized — so its discovery proves all the more dramatic.

In most accounts the Jesuits figure prominently as the scientific adversaries of Galileo in the two decades prior to his trial. Orazio Grassi, the leading Jesuit scientist, attacked Galileo in print, a compliment which was returned ten-fold and

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

from which Grassi was never to recover. Christopher Scheiner, a brilliant Jesuit astronomer, also suffered severely in debate with Galileo. But the precise role of the Jesuits behind the scenes has inevitably remained obscure, partly it is claimed because of their deviousness and partly because of the secrecy of the inquisition. Though there exists circumstantial evidence to support such views they have always been strenuously denied by a well-developed tradition of Jesuit apologetic historiography.

The situation has been transformed as a result of Redondi's research. This book should therefore prove compulsive reading for anyone with an interest in the emergence of modern science. It is one of the most interesting and important works on Galileo's relationship with the church for decades.

The stumbling block for Galileo was not his ego, nor Urban VIII, nor, according to Redondi, even Copernicanism, but the post-Tredentine interpretation of the Eucharist. It was this interpretation that had established an immovable barrier between Catholic and Protestant. The significance of Catholic Eucharistic dogma was therefore paramount from the viewpoint of the combative counter-reformation theology of the Jesuits. It had to be protected with the utmost vigilance against all comers, and at all costs.

A particular strength of Redondi's work is that from his analysis of a wealth of writings by Jesuits he reveals the full extent of the conflict generated between Eucharistic dogma and the rising tide of atomism. This was unquestionably an issue of great moment for seventeenth-century science. That he has demonstrated

Mary Evans

this to be a continuing and overriding priority of the Jesuit scientists provides an important new perspective. For, as Redondi shows, this conflict surfaces again and again, and was instrumental in dissuading Descartes in 1636 from proceeding with his plans for publishing his atomistic cosmology in *Le Monde*. But although Redondi convincingly demonstrates the relevance of the development of the Jesuit programme on a broad front, it is not as easy to substantiate the specific and very bold claim that it was the issue of Eucharistic dogma, rather than Copernicanism, that was at stake in Galileo's trial.

The historical source of Redondi's thesis is well known in outline — the bad relations that developed between the Jesuits of the Collegio Romano in the period following Galileo's open attack on Aristotelianism, beginning with his *Discourse on Bodies on Water* in 1612 and exacerbated by his priority dispute with Scheiner over the discovery of sun-spots. The *Discourse* reveals the first signs of Galileo's commitment to atomism — a pernicious doctrine from the viewpoint of post-Tredentine Dogma, for it asserted that the qualities of taste, colour and so on were subjective impressions produced in the perceiver and not in reality part of the object of perception. Only the primary qualities of extension and motion had claim to objective existence; other qualities, according to Galileo, were merely secondary.

A crucial consequence of the Council of Trent had been the establishment as dogma that in the Eucharist the host became in reality the body and blood of

Christ, though it retained the outward appearance of bread. The alliance of Christianity with Aristotelian philosophy that produced scholasticism provided an epistemology that underpinned this view of the Eucharist. In scholasticism it was possible for qualities to exist without a subject. It therefore appeared to Jesuits, and others (though certainly not to all theologians), that atomism represented a denial of this view of the Eucharist, and was in short heretical.

Galileo's most outspoken expression of his atomistic epistemology appeared in *The Assayer* in 1623, a work in which he attacked the Aristotelianism of Orazio Grassi's earlier book *The Astronomical and Philosophical Balance*, while advancing his own view of scientific method. It was hailed as a triumph by his contemporaries, a masterpiece both of style and argument that had effectively demolished his Jesuit adversary. Having lost their battle on the scientific front, Galileo's enemies turned to theology just as they had done a decade earlier in the dispute

over Copernicanism. Redondi has discovered in the Vatican files a condemnation of the *Assayer* submitted to the inquisition in 1624. It accused the *Assayer* of contradicting the Eucharistic dogma. The legal process that should have ensued was however mysteriously blocked immediately afterwards by the intervention of a powerful prelate. But again, as with other aspects of the affair, no explanation of this event is available. On the other hand, Redondi has gathered considerable evidence to suggest the condemnation could have been a time bomb quietly ticking in the background awaiting its subsequent explosion in 1632.

Was this then the real issue at stake in Galileo's trial? Redondi has uncovered much fascinating circumstantial evidence suggesting that it could have been — but nothing to show that it was ultimately the issue that brought Galileo to trial. There is nevertheless a vast amount of evidence indicating that it was Copernicanism that caused his downfall. The proposal that after all the trial bore no relation to the

legal case as presented, and in reality concerned an entirely different and distinct heresy, does of course carry the conspiracy theory literally to incredible lengths.

There is, over and above this, one apparently insurmountable difficulty for Redondi's thesis, even if we were inclined to ascribe it a prominent rather than predominant role in the events of 1633. If we were to assume that Eucharistic dogma was at the heart of the trial and had resulted in the anathematizing of Galilean atomic theory — albeit secretly — how are we to explain Galileo's most important explication and application of the theory in his *Discourse on the Two New Sciences* that was published five years later? The penalty for a relapsed heretic was death, as everyone knew. Galileo had been extremely frightened in 1633 — he would have been terrified of a further confrontation. □

Ron Naylor is Head of Philosophy at Thames Polytechnic, Wellington Street, London SE18 6PF, UK.

ADVERTISEMENT

Smithsonian Ethnographic Inquiry Series

Series Editors: William L. Merrill, Ivan Karp

PILGRIMS OF THE ANDES

Regional Cults in Cusco

Michael J. Sallnow, *London School of Economics*

An in-depth study, of the Quechua-speaking peasants of The Southern Peruvian Andes, their land, culture, history and Church.

1987, 330pp, 29 b & w illus., Cloth £21.95

LEGENDS OF PEOPLE, MYTHS OF STATE

Violence, Intolerance, and Political Culture in Sri Lanka and Australia

Bruce Kapferer, *University College, London*

In this fascinating book, Bruce Kapferer inquires into the ideological foundations of modern political cultures.

1987, 280pp, 20 b & w illus., Cloth £21.95, Paper £10.95

Smithsonian Institution Press 3, Henrietta Street, Covent Garden, London WC2E 8LU or telephone 01-240 0856.

Reader Service No.7

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

Universal appeal

David W. Hughes

Exploring the Southern Sky: A Pictorial Atlas from the European Southern Observatory (ESO). By S. Laustsen, C. Madsen and R.M. West. *Springer-Verlag: 1987. Pp. 274. DM98 (DM128 in 1988), £36.**

NEARLY half of the world's 6,000 professional astronomers work in Europe, and so can view only the northern sky. To redress the balance, in October 1962 Belgium, France, the Netherlands, Sweden and West Germany combined together to build and run the European Southern Observatory (Denmark joined in 1967 and Italy and Switzerland in 1982).

The site they chose was a 2,400-m-high mountain ridge, officially known as Cerro Chincado, but colloquially called La Silla (The Saddle). The observatory is some 600 km north of Santiago in Chile. To the east is the massive Andes mountain range, to the north the Atacama desert. Seventy kilometres to the west is the Pacific Ocean with its cold Humboldt Current. The average rainfall is 5 cm per year. During the night the temperature changes by less than 3 °C. The sky is clear for more than six consecutive hours on more than 70 per cent of all nights. It is one of the finest astronomical observing sites on Earth.

Many observatories have been in existence for over 25 years, but few have celebrated this anniversary in such a splendid way as the ESO. Laustsen, Madsen and

West have collected together 90 colour plates and 147 black-and-white prints, all taken at the observatory during the past ten years, and have produced a superb picture-book of southern celestial objects and star fields. It is the best such book I have seen. Authors, publishers and printer have excelled themselves.

Of course, the usual favourites are here — the Magellanic Clouds, galaxies such as the Sombrero, the Sculptor Group, NGC 253, NGC 2997, open clusters like NGC 6193 in Ara, the Orion Nebula, the Vela supernova remnant, the Coal Sack, Eta Carinae, globular cluster M55 and Omega Centauri. Nearby objects are not forgotten. There is a section on comets and asteroids plus a chapter on the observatory itself, with aerial photographs of the complex and closeups of the telescopes. There are also rarer sights, such as minor-planet trails near the ecliptic, the asteroid La Silla, SN 1987 A, and an amazing objective prism Schmidt image of the star field near M17.

Each picture has a detailed caption. Not only is there a discussion of the object but we are also told which filters, detectors and emulsions were used, the exposure time and the image scale of the printed plate. The book is more than a joy to look at, it is a useful tool for both amateur and professional astronomers. As a bonus there are 30 maps adjacent to the plates, so that readers can find their way in amongst the stars and nebulae, and also a 120 cm foldout panorama of the Milky Way. Altogether this is an astronomical *tour de force*. □

David W. Hughes is Senior Lecturer in Astronomy and Physics, University of Sheffield, Sheffield S3 7RH, UK.

*A German edition of the book, published by Birkhäuser, Basel, is available in West Germany, Austria and Switzerland.

Romer renewed in the 1980s

Alec Panchen

Vertebrate Paleontology and Evolution.
By Robert L. Carroll. W.H. Freeman:
1987. Pp.698. \$52.95, £47.50.

A LITTLE over two decades ago there appeared the third and last edition of Alfred Sherwood Romer's *Vertebrate Paleontology*. This was taken to be the textbook on the subject by all students in the English-speaking world.

It is inevitable that Carroll's new text will not only be compared with Romer's, but will be regarded by many as the fourth edition. Although produced by different publishers, the physical resemblance between the two books is such that it is difficult to see it as coincidence. Both are handsomely produced quarto volumes printed in double columns and profusely illustrated with line drawings of reconstructed fossils but with very few pictures of specimens. Both incorporate an appendix consisting of an attempt to list all genera of fossil vertebrates in a traditional classification of the group.

More important, however, is the similarity of approach. It was widely accepted until the late 1960s that the tasks of a vertebrate palaeontologist were to describe and restore the anatomy of fossil vertebrates from collected specimens, to attempt some reconstruction of their mode of life with a particular emphasis on functional morphology, and to use details of the fossil record to demonstrate, not just the truth of evolution, but fragments of its pattern. Finally, but often as an afterthought, the working palaeontologist felt called upon to fit his specimens into a received pattern of classification. Few, however, questioned the largely intuitive nature of their classificatory method.

Carroll has had a distinguished career working within this traditional pattern and his book reflects this. Like Romer's it is almost entirely concerned with a systematic treatment of fossil vertebrates. After an introductory chapter and some account of theories of the origin of vertebrates, most of the text is taken up with a group-by-group review of vertebrate fossils, beginning with the most primitive fish-like forms and finishing with mammals. The account of each group includes anatomical descriptions, along with something on functional morphology, relationships and reconstructed evolution. Unlike Romer, Carroll notes controversies in the text with citations of the literature. References cited in each chapter are listed at its end, probably the most convenient method in a textbook.

I endorse the carefully chosen words of

a colleague, quoted on the dust jacket: "I am greatly impressed by the data presented. . . . Useful not only for students but for colleagues in the field as a resource"; and yet I know that many other people will not just be dissatisfied, but actually annoyed with Carroll's book. This is not simply because they will find errors in their own field of expertise. These are inevitable in a large text of such wide coverage, and the present work is by no means free of them. It is the whole approach to the subject that many will object to. The critics will, I suggest, be divided into two, not mutually exclusive, categories.

The first group will point to Carroll's lack of rigour in matters of classification. The publication of the last edition of Romer coincided almost exactly with that of Hennig's *Phylogenetic Systematics*. Hennig advocated a rigorous method of classification in which species and higher taxa were grouped exclusively by the use of uniquely shared characters. The results of such grouping are expressed in branching diagrams known as cladograms before being turned into orthodox classifications or interpreted as phylogeny, the pattern of evolution. Most vertebrate palaeontologists now agree that cladistics represents a methodology of classification which fills what was almost a vacuum amongst the techniques of their subject. Carroll gives quite a lengthy account of classification in general and cladistics in particular in his first chapter, but his heart appears not to be in it, so that in the body of the book there is a glorious jumble of approaches to classification and phylogeny and an amazing assortment of diagrams.

The second category of objectors would demand an entirely different sort of text, guided by the question, "What is vertebrate palaeontology for?". Carroll does

consider the general interpretation of the fossil record in his final chapter on evolution, and also has a chapter in the body of the book on the biology and extinction of dinosaurs. His views on evolution and extinction are orthodox 'neo-Darwinian'. Critics would argue, however, that a text on vertebrate palaeontology should be organized to demonstrate the use that can be made of the fossil record and the practical and theoretical problems tackled by palaeontologists from the discovery and interpretation of fossils to their use in evolutionary and other hypotheses — a topic — rather than a taxon-based book.

Such a topic-based book, Olson's *Vertebrate Paleozoology*, was published soon after Romer's last edition, but is now out of print. Jarvik's scholarly but idiosyncratic *Basic Structure and Evolution of Vertebrates* (1980) in a sense combines both approaches but deals almost entirely with fish-like forms and is limited to comparative anatomy. The ideal topic approach, though not specifically vertebrate, was Raup and Stanley's *Principles of Paleontology* (2nd Edn, 1978) which cries out for a third edition. Perhaps one needs both a taxon-based text, such as Carroll's, and a topic-based one, but it is unlikely that any but the most devoted student would buy both.

I do not want to end this review on a negative note. Carroll has to his credit an immense amount of useful labour in writing the book and will probably corner the market for a vertebrate palaeontology text for the rest of this century. I am glad to have the book and anticipate using it frequently. I suspect that will also be the case for many of his critics. □

Alec Panchen is Reader in Vertebrate Zoology, Department of Zoology, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK.



Mammoth weapon — part of a weighted spear thrower, broken through the shaft below the carving of a mammoth which forms the weight of the thrower. The tail of the animal originally formed the hook, and the line of its back follows the line of the shaft. The picture is taken from A Catalogue of Palaeolithic Art in the British Museum by Ann Sieveking. The collection comes mainly from France, and while the greater part was acquired before 1890, it has not previously been described in detail. The book is published by the British Museum and costs £70.

Cellular control

Sydney Shall

ADP-Ribosylation of Proteins: Enzymology and Biological Significance. By Felix R. Althaus and Christoph Richter. Springer-Verlag: 1987. Pp.237. DM198.

CONTROL of cellular reactions is often realized by the covalent modification of previously synthesized enzymes, receptors, structural proteins and so on. The covalent ADP-ribosylation of proteins is one such regulatory mechanism.

ADP-ribosylation may be either monomeric or polymeric. Mono-ADP-ribosyl transferases include an array of bacterial toxins and a number of endogenous cellular enzymes. The bacterial toxins are the easiest to study — we know more about them than about the cell enzymes — and are described in the first part of this book. Diphtheria, *Pseudomonas*, cholera, pertussis (whooping cough), *E. coli* and some minor botulinum toxins all consist in part of ADP-ribosyl transferase enzymes. These toxins (enzymes) transfer ADP-ribose from NAD to a very specific amino acid residue in a very specific target protein — all of which are apparently so-called G-proteins or GTP-binding proteins. Because endogenous enzymes with similar specificities have been described it is likely that ADP-ribosylation will emerge as an important mode of regulation of many G-proteins.

The other part of the book deals with poly-ADP-ribosylation reactions on chromatin proteins. Felix Althaus has a strong prejudice in favour of the notion that poly-ADP-ribosylation reactions modulate chromatin conformation. He has argued this idea frequently and with great conviction, and does so again here. However, although I found the exercise of preferences and prejudices most admirable and stimulating, and some elegant experiments with purified components support the notion, the book describes no convincing observations which show that poly-ADP-ribosylation reactions modify chromatin structure in intact cells. Of course, poly-(ADP-ribose) is a large polyanion, and this implies that such a polymer is very likely to modify chromatin structure. In my view, this last argument is at present the best one in support of Althaus's views.

The authors also discuss the demonstration, some years ago, that poly-ADP-ribosylation reactions are an integral part of the cellular response to DNA-damaging agents, probably in direct proportion to the steady-state level of DNA nicks and breaks, and probably also only in a regulatory not in a mechanistic role. The practical therapeutic potential of these observations is currently being explored in a

number of laboratories, but the precise molecular explanation of the function of poly-(ADP-ribose) in DNA repair is still being defined. My colleagues and I have published some evidence indicating that perhaps poly-(ADP-ribose) polymerase (indirectly?) modulates the activity of DNA ligase 11. For the most part, this history is accurately and clearly described in the book. However, it seems to me that the authors have not discriminated sufficiently clearly between data derived from intact cells (or organisms) and those derived from isolated or purified components. Although both types of data are valid, they are not at all equivalent.

A comprehensive monograph takes time to produce and the field has not waited upon the publishers; the flowers

have continued to bloom. Thus, the molecular biology of poly-(ADP-ribose) polymerase is not recorded in this book; the gene has been cloned in at least six laboratories during 1987. Also, the work on botulinum toxins C and D is not described. These omissions merely indicate the rapid pace of advance; the number of publications in this field has been growing exponentially since 1980.

Althaus and Richter have produced a well-written, comprehensive and stimulating review of this rapidly expanding area of research. Although it must be read critically, the book will be the standard account of ADP-ribosylation of proteins for many years to come. □

Sydney Shall is a Professor in the School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK.

Never spare the rod

David W. Kingsbury

The Rhabdoviruses. Edited by Robert R. Wagner. Plenum: 1987. Pp.544. \$89.50 (North America), \$107.40 (elsewhere)

FOR biological mystery and glamour, consider the rhabdoviruses. This large family of ultramicroscopic rodlets (*rhabdos*, rod) has evidently been a companion and occasional pathogen of multicellular organisms for eons, considering the impressive variety of phyla afflicted by its members. A long evolutionary history is indicated by the existence of rhabdoviruses that infect plants (including food staples), insects (such as the hereditary sigma virus of *Drosophila*), cold-blooded vertebrates (salmon and trout among them) and warm-blooded vertebrates (vesicular stomatitis virus of cattle and swine has a significant impact on human affairs).

The most serious human pathogen among the rhabdoviruses is the causative agent of rabies. Louis Pasteur tamed *la rage* by selecting avirulent rabies virus mutants and injecting them into infected subjects. Luckily for him, for those injectees and for the rest of us, the long incubation period of the disease opens a sufficiently wide window of time for the mutant viruses to stimulate curative antibodies. Pasteur's achievement was a cornerstone of the science of immunology, and it established a precedent for the development of attenuated vaccines prophylactic against many other viruses and bacteria.

Rhabdoviruses seem simple genetically, possessing only five or six genes in a single strand of RNA. Adding complexity, the RNA strand is complementary to the messenger RNAs that specify viral proteins, placing the rhabdoviruses in the broad category of negative-stranded viruses. This *modus vivendi* entails the

dedication of precious genetic information to specifying enzymatic proteins that can transcribe and replicate the viral genome. But transcription offers a versatile means of regulating gene expression, a benefit that has evidently been worth the price, given the evolutionary success of negative-strand viruses. Despite much progress, many important questions about their replication are unanswered, because they do not engage in genetic recombination, nor are they easily manipulated biochemically or by recombinant DNA technology.

The Rhabdoviruses is the latest in a series, *The Viruses*, comprising volumes of reviews devoted to individual virus families, edited by H. Fraenkel-Conrat and R.R. Wagner. This volume maintains the high standards of the series, which updates and expands upon an earlier set, *Comprehensive Virology*, a product of the same editorial team.

Wagner has assembled 11 separately authored reviews that assess progress in the field, giving due weight to controversies and unsolved problems. The editor himself contributes two overview chapters. He and his coauthors are all well-known authorities, actively engaged in research on their topics, and are skilful communicators. Most of the chapters deal with the molecular genetics of the popular experimental model of the family, vesicular stomatitis virus (VSV), but there is a chapter on rabies virus pathogenesis, one on plant rhabdoviruses, and a brief review of the ecology of rhabdoviruses that infect vertebrates. In addition, the chapters on VSV are enriched and balanced by abundant references to related work on other members of the family, making the book valuable not only to specialists but also to a broader audience. □

David W. Kingsbury is in the Department of Virology and Molecular Biology, St Jude Children's Research Hospital, PO Box 318, Memphis, Tennessee 38101, USA.

Entropy and cosmology

Edgard Gunzig, Jules G  heniau & Ilya Prigogine*

Facult   des Sciences (CP 231), Universit   Libre de Bruxelles, B-1050 Bruxelles, Belgium

A cosmological model is proposed in which an inflationary de Sitter spacetime appears as the result of a fluctuation in the conformal degree of freedom of an initial Minkowski vacuum. A population of black holes is thereby created, which evaporate during the inflationary phase. Then a second phase transition turns the de Sitter cosmology into the usual Robertson-Walker universe. The temperature and specific entropy per baryon of the present universe are deduced, and depend only on the mass of the black holes and on the universal constants h , c and k .

ANY spacetime obeying the cosmological principle, that the Universe is spatially homogeneous and isotropic, must have a metric of the Robertson-Walker (RW) form, with line element $ds^2 = dt^2 - R^2(t)dl^2$, where t is the cosmological proper time, $R(t)$ is the scale-factor and dl^2 is the line element for a three-dimensional space of constant positive, negative or zero curvature. Any RW metric is in turn conformally related to the Minkowski metric by $ds^2 = F^2 \eta_{\mu\nu} dx^\mu dx^\nu$, where $\eta_{\mu\nu} = \text{diag}(1, -1, -1, -1)$ represents the Minkowski metric. If, as here, we consider only spatially open (infinite) spaces, the conformal factor F is a function only of the Lorentz invariant $x = (\eta_{\mu\nu} x^\mu x^\nu)^{1/2}$ and in addition¹ $R = xF$ and $dt = F dx$. In the conformal system, Einstein's equations become

$$R_\nu^\mu - \frac{1}{2} \delta_\nu^\mu R = F^{-4} \left(2F\ddot{F} - 4\dot{F}^2 - \frac{2}{x} F\dot{F} \right) u_\nu u^\mu - \left(2F\ddot{F} - \dot{F}^2 + \frac{4}{x} F\dot{F} \right) \delta_\nu^\mu = -\kappa T_\nu^\mu \quad (1)$$

where $u^\mu = x^\mu/x$, $u_\mu = \eta_{\mu\nu} x^\nu$, a dot denotes differentiation with respect to x and $\kappa = 8\pi G$ is the gravitational constant. (In this equation, R is the Ricci tensor or scalar.) The energy-momentum source in this equation then has the form

$$T_\nu^\mu = (\rho + p)u^\mu u_\nu - p\delta_\nu^\mu \quad (2)$$

which reduces Einstein's equations to the two coupled equations

$$2F\ddot{F} - 4\dot{F}^2 - \frac{2}{x} F\dot{F} = -\kappa(\rho + p)F^4 \quad (3)$$

and

$$2F\ddot{F} - \dot{F}^2 - \frac{4}{x} F\dot{F} = -\kappa pF^4 \quad (4)$$

where ρ and p are functions of x only.

Recently, the possible dynamical role of the conformal degree of freedom has received some attention²⁻⁵. It can cause an instability of the Minkowskian vacuum⁶⁻⁸, leading to the creation of particles with mass about fifty times the Planck mass M_{Pl} , which are interpreted as small black holes, and can also mediate an irreversible transformation of gravitational energy into matter^{9,10}. From these starting points, we construct a more detailed description of cosmological evolution.

The conformal factor, F , or equivalently the determinant of the metric $g = F^8$, is now taken as a dynamical degree of freedom. The energy associated with it is unbounded from below²⁻⁵, and

it is this property which confers on it the role of a reservoir of negative energy from which the energy to create the black hole population is extracted. The cosmological constituents are thus created at the expense of curving spacetime, and vice versa; the positive energy carried by the first is compensated by the negative energy of the second^{2-5,11-14}. We call such a process, and the resulting cosmology, a self-consistent one.

In contrast to fluctuations in the conformal degree of freedom, deviations from purely conformal Minkowski metrics (that is, inhomogeneities in spacetime) make a positive contribution to the energy. Such fluctuations are associated with the propagation of gravitational waves^{15,16}. The energy contribution can be seen by considering a metric conformally related to a perturbed Minkowskian background, so that

$$g_{\alpha\beta} = l^2 \Phi^2 (\eta_{\alpha\beta} + h_{\alpha\beta}) \quad (5)$$

where l is an arbitrary lengthscale. Provided suitable gauge conditions are imposed on the fluctuations $h_{\alpha\beta}$, the general relativistic action becomes

$$-\frac{1}{2} \kappa \int \sqrt{-g} R d^4x = -\frac{1}{2} \int d^4x \left(\eta^{\alpha\beta} \Phi_\alpha \Phi_\beta - \frac{\Phi}{24} h_{\alpha\beta}^{\mu\nu} h_{\mu\nu}^\alpha \right) \quad (6)$$

where Φ_α and $h_{\alpha\beta}^{\mu\nu}$ are ordinary derivatives. The conformal and local degrees of gravitational freedom thus contribute with opposite signs.

The energy transfer from gravitational curvature to matter is taken to be an irreversible process leading to a burst of entropy associated with the creation of matter^{9,10}. As far as energy is concerned, the creation of the universe from a Minkowskian vacuum is indeed a 'free lunch'. Energy does not distinguish between the initial vacuum and the de Sitter cosmology. The distinction is provided by the creation of entropy, which makes the primeval instability an irreversible process on the cosmic scale. (This requires a reinterpretation of the meaning of the energy density and pressure in the presence of matter creation. This will be discussed in a separate paper¹⁰.)

In the cosmological history described here, the universe emerges not from an initial singularity but from a vacuum instability resulting in the creation of masses greater than a threshold mass $m_{\text{th}} = (288\pi^2)^{1/2} m_{\text{Pl}}$. As shown in earlier work^{2-4,17}, this leads to an inflationary de Sitter universe in which the constant energy density is related to the mass of the created particles by

$$\rho_d = 30(\kappa m_{\text{th}}^2) \kappa^{-2} \varepsilon; \quad \varepsilon = \left(\frac{m}{m_{\text{th}}} \right)^2 - 1 > 0 \quad (7)$$

The order of magnitude of m leads one to think of the created masses as small black holes^{2-4,18}.

In what follows we use units with $\hbar = c = k = 1$, in which the gravitational coupling constant is $\kappa = 6.56 \times 10^{-65} \text{ cm}^2$, the proton mass is $m_p = 4.75 \times 10^{13} \text{ cm}^{-1}$ and $1\text{K} = 4.36 \text{ cm}^{-1}$. The essential characteristics of a black hole of mass M are¹⁹ its

* Present address: Center for Studies in Statistical Mechanics and Thermodynamics (RLM-7), University of Texas at Austin, Austin, Texas 78712, USA.

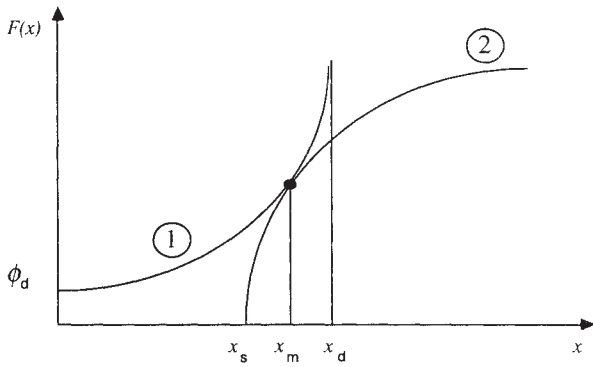


Fig. 1 The continuous matching between the two conformal factors. Curve (1) represents the de Sitter expansion and curve (2) the usual matter-radiation Robertson-Walker evolution. The singularity (outside the physical range in the present model) occurs at conformal time x_s ; x_m is the matching time between both cosmological regimes, and x_d denotes the asymptotic conformal de Sitter time. Their ratios are defined by $x_m = \Delta_1^{1/2} x_d$, ($\Delta_1 < 1$) and $x_m = \Delta_2 x_s$, ($\Delta_2 > 1$). For clarity, this figure obviously does not respect the relative scales of the various typical points.

temperature, $T = (\kappa M)^{-1}$, its entropy, $S = \kappa M^2/2$ and its life-time against quantum evaporation, $t_e = (640/27\pi)(\kappa^2 M^3/N)$, where N is the total number of helicity states of massless and massive particles emitted by the black hole²⁰.

Robertson-Walker and de Sitter cosmologies

The conformal factor for the spatially open de Sitter metric is

$$F_d(x) = \phi_d(1 - x^2/x_d^2)^{-1} \quad (8)$$

where x , defined above, is used as the conformal time coordinate and ϕ_d is an arbitrary amplitude. The function $F_d(x)$ is the solution of Einstein's equations (3) and (4) for pressure and density $-p = \rho = \rho_d = \text{constant}$. The integration constants ϕ_d and x_d are related to ρ_d by $12(\phi_d x_d)^{-2} = \kappa \rho_d$. Cosmological proper time t is related to conformal time by

$$t = \int_0^x F_d dx = \frac{1}{2\sqrt{\mu}} \ln \left(\frac{1+x/x_d}{1-x/x_d} \right) \quad (9)$$

where $\mu = (x_d \phi_d)^{-2}$. The conformal factor F_d is expressed in terms of t as

$$F_d(t) = \frac{\phi_d(\exp(2\sqrt{\mu}t) + 1)^2}{4 \exp(2\sqrt{\mu}t)}, \quad (10)$$

and $R_d(t) = x(t)F_d(t)$. Conformal time in the range $0 < x < x_d$ corresponds to the physically allowed range of cosmological time $0 < t < \infty$.

The conformal factor F_{mr} for an RW universe containing matter and radiation may be written

$$F_{mr} = \alpha_+ x_s^{-1} - \alpha x^{-1} + \alpha_- x_s x^{-2} \quad (11)$$

This is the solution of equations (3) and (4) with $\rho = \rho_m + \rho_\gamma$ where ρ_m and $\rho_\gamma = 3p_\gamma$ are the energy densities of matter and radiation respectively. The constants α , α_+ and α_- are defined by $2\alpha_\pm = \alpha \pm \beta^{1/2}$, where $6\alpha = \kappa \rho_m R^3$ and $3\beta = \kappa \rho_\gamma R^4$. The singularity in this model occurs at conformal time $x = x_s$, because $F_{mr}(x_s) = 0$. This corresponds to proper time $t = 0$, although this point lies outside the physical range of the cosmological model presented here (see Fig. 1).

The Hubble function is given by

$$H = \frac{1}{R} \frac{dR}{dt} = \frac{\dot{R}}{R} \frac{dx}{dt} = \frac{\dot{R}}{RF} \quad (12)$$

where a dot denotes derivation with respect to conformal time x . From equation (11) we derive the useful relation $HR^2 =$

$\alpha_+ x/x_s - \alpha x_s/x$, so that $H^2 R^4 - R^2 = \beta + 2\alpha R$. The definitions of α and β then lead to

$$R^2 H^2 (1 - \frac{1}{3} \kappa \rho_m H^{-2} - \frac{1}{3} \kappa \rho_\gamma H^{-2}) = 1 \quad (13)$$

Ensuring that Einstein's equations are obeyed at conformal time x_m , when the de Sitter and RW regimes are joined, is the same as ensuring that the conformal factors and the derivatives with respect to conformal time are equal. This also ensures continuity in the energy density. If we introduce the time ratios $x_m/x_d = \Delta_1^{1/2}$ and $x_m/x_s = \Delta_2$ (where, obviously, $\Delta_1 < 1$ and $\Delta_2 > 1$), then the matching conditions may be written

$$\alpha(\Delta_2 - 1) + \beta^{1/2} = C \Delta_2 \quad (14)$$

and

$$\alpha \Delta_1 (\Delta_2 - 1)^2 + \beta^{1/2} (\Delta_2^2 - 1) \Delta_1 = C(1 - \Delta_1) \Delta_2 \quad (15)$$

where $C = 2\Delta_1^{3/2}/\sqrt{\mu}(1 - \Delta_1)^2$. From all the definitions in this section, we can now retrace cosmological evolution backwards from the present day. The matching conditions then impose constraints on the properties of the constant density medium of the de Sitter phase.

Constraints on de Sitter cosmology

The present temperature of the cosmological microwave background radiation is $T_0 = 11.8 \text{ cm}^{-1}$, corresponding to a radiation energy density $\rho_{\gamma 0} = a T_0^4 = 1.28 \times 10^4 \text{ cm}^{-4}$, where the constant $a = \pi^2/15$ in our units. The density of matter may be written $\rho_{m0} = 10^7 v \text{ cm}^{-4}$, where the factor v is to accommodate observational uncertainty and has a value greater than 1.3. If the RW universe is to be spatially open, then $\kappa \rho_{m0} < 3H_0^2$, where the numerical value of the Hubble constant is $H_0^{-1} = 10^{28} u \text{ cm}$, where the factor u includes observational uncertainty. We take $1 < u < 1.8$, and obtain $2.18 vu^2 < 100$. The numerical values for matter and radiation density and for the Hubble constant deliver three equations for the three unknowns α , β and x_0/x_s . From equation (13), the present value of the RW scale factor is $R_0 = 10^{28} \tilde{u} \text{ cm}$, where

$$\tilde{u} = u(1 - \frac{1}{3} \kappa \rho_{m0} H_0^{-2} - \frac{1}{3} \kappa \rho_{\gamma 0} H_0^{-2})^{-1/2} \quad (16)$$

The invariant $RT = 1.18 \times 10^{29} \tilde{u}$, leading to $\alpha = 1.09 \times 10^{26} \tilde{u}^3 v \text{ cm}$ and $\beta = 0.28 \times 10^{52} \tilde{u}^4 \text{ cm}^2$. The present density ratio of matter and radiation is then $\rho_{\gamma 0}/\rho_{m0} = 1.28 \times 10^{-3}/v$, giving an adiabatic invariant $(\rho_{\gamma 0}/\rho_{m0})/T_0 = 1.08 \times 10^{-4}/v$, or a specific entropy per proton²¹

$$s = \frac{n_{\gamma 0}}{n_{m0}} = 3.7 \frac{\rho_{\gamma 0}}{\rho_{m0}} \cdot \frac{m_{pr}}{T_0} = 1.9 \times 10^{10}/v \quad (17)$$

where $n_{\gamma 0}$ and n_{m0} are the photon and proton number densities respectively. It should be noted that values $u = u_{\min} \approx 1$ and $v = v_{\min} \approx 1.3$ correspond to $\tilde{u} = 1$.

The temperature at the matching point is deduced by assuming that the de Sitter energy density remains constant during the de Sitter regime, and that the same is also true for the temperature. Moreover, this temperature is identified with the initial black hole temperature. It will be shown in the next section that the initial mass of the black holes differs infinitesimally from the threshold mass m_{th} corresponding to instability. The needed temperature is therefore given by $T_d(x_m) = 1/\kappa M$, with $\kappa M^2 = 288\pi^2$, and if continuity is assumed the temperature at the beginning of the RW regime is the same, namely $T_{mr}(x_m) = 2.3 \times 10^{30} \text{ cm}^{-1}$. The photon energy density is then $\kappa \rho_\gamma(x_m) = 2.4 \times 10^{56} \text{ cm}^{-2}$. The same continuity requirements mean that this is also the constant energy density of the de Sitter regime, neglecting the matter contribution. We then find in turn $x_d \phi_d = 10^{28} \text{ cm}$ for the constants defined in equation (8). The RW scale factor at the matching time is $R(x_m) = 5 \times 10^{-2} \tilde{u} \text{ cm}$ and the matching time itself is $t_m = (3.11 + \ln(2\tilde{u})/20) \times 10^{-27} \text{ cm}$. According to the value of $\tilde{u}$, this time is unbounded from above; this reflects the fact that an infinite value of $\tilde{u}$ corresponds to

a spatially flat universe, for which a different coordinate parametrization is needed. The lower bound, corresponding to $\tilde{u} = 1$, is given by $t_m(\min) = 3.1 \times 10^{27}$ cm. It is intriguing that this lower limit for the matching time is exactly the same as the evaporation time of black holes whose initial mass is the instability threshold mass, provided that the number of species emitted during evaporation is $N = 3$. This corresponds, for example, to a medium in which there is only one type of scalar particle in addition to photons. Thus there is consistency with a cosmological model in which instability leads to the creation of a population of primordial black holes.

Constraints on Robertson–Walker cosmology

A basic feature of the de Sitter cosmology is the constancy of the physical (and geometrical entities) such as the energy density ρ_d . The de Sitter temperature T_d and the specific entropy s_d are also taken to be constant. We now show that if the de Sitter phase is constituted of black holes, the number of unknowns can be reduced from four to two. Identification of the de Sitter temperature with that of the black holes at the onset of their evaporation, together with the continuity of the energy density and temperature at the matching point, imply the relation $\rho_d = a(\kappa M)^{-4}$. This, applied directly to equation (7), shows that the parameter $\varepsilon = 10^{-12}$, which means that the mass of the black holes differs only infinitesimally from the threshold mass, M_{th} . The energy density is thus $\rho_d = a(\kappa M_{th})^{-4}$, and the parameter μ of equation (9), relating conformal and proper times is $\sqrt{\mu} = (a/12)^{1/2} \kappa^{-3/2} M_{th}^{-2} = 10^{28}$ cm $^{-1}$. Furthermore, as mentioned in the previous section, the black hole evaporation time t_e corresponding to three degrees of freedom is strictly equal to the minimum matching time t_m defined at the end of the previous section. All the numerical results of the previous section will therefore be recovered in our ‘toy’ model, which involves only massive scalar particles in addition to photons.

It is worth demonstrating that the generalized Second Law of Thermodynamics, as applied to black holes¹⁹, imposes an upper bound on the temperature at the matching point and thus on the initial RW matter-radiation temperature, which is the same as the constant temperature during the de Sitter phase. The temperature as well as the energy density is continuous at the matching point. As applied here, the generalized second law requires that the de Sitter entropy density s_d has to be lower than or equal to the resulting RW matter-radiation entropy density s_{mr} evaluated at the matching point. During the de Sitter phase, s_d is constant and is most easily evaluated at the very beginning of the phase, as evaporation begins. The entropy of an individual black hole of mass M_{th} is $S_{bh} = \kappa M_{th}^2/2$, and their number density is $n_d = \rho_d/M_{th}$, so $s_d = \rho_d \kappa M_{th}/2$. Moreover, $s_{mr} = (\rho_{mr} + p_{mr})/T = 4\rho_d/3T$, so that the requirement $s_{mr} \geq s_d$ implies that $T \leq \frac{8}{3}(\kappa M_{th})^{-1}$, or equivalently $\rho_d < (8/3)^4 a T_{bh}^4$.

The present model is only appropriate if the black hole fluid is dilute, and would clearly fail if black holes overlapped or were close enough that they would coalesce into a smaller number of larger black holes. The Schwarzschild radius is initially $R_s = \kappa M_{th}$, so their ‘close-packing’ number density is $n_{cp} = 3/(4\pi \kappa^3 M_{th}^3)$ corresponding to an energy density $\rho_{cp} = 3/(4\pi \kappa^3 M_{th}^2)$. This is higher than the actual density $a(\kappa M_{th})^{-4}$ by $\sim 10^3$, so the black holes do indeed constitute a dilute gas.

We can now follow the forward evolution of our model from the creation of the black holes to the present day. The exponent in the de Sitter conformal function (10) is

$$\sqrt{\mu} t_e = \left(\frac{a}{12}\right)^{1/2} \frac{640}{81\pi} (\kappa M_{th}^2)^{1/2} = 31.5 \quad (18)$$

The matching point is located in the large-time asymptotic regime of the de Sitter phase, and the scale-factor simplifies to

$$R(t_e) = \frac{1}{4\sqrt{\mu}} e^{2\sqrt{\mu} t_e} = 5.1 \times 10^{-2} \text{ cm} \quad (19)$$

This is obviously very sensitive to the value of $\sqrt{\mu} t_e$; even a

slight modification of its value would drastically alter the value of $R(t_e)$. We postulate also that the constant de Sitter temperature is given by the black hole temperature at their creation, $T(x_m) = 2.3 \times 10^{30}$ cm $^{-1}$. The quantity RT at the matching point is thus $R(x_m)T(x_m) = 1.18 \times 10^{29}$, and this is also the value of the matter-radiation RW invariant. The parameter β of the RW evolution equation (11) is then $\beta = \kappa a(RT)^4/3 = 0.27 \times 10^{52}$ cm 2 . The evaporation time t_e fixes the conformal matching factor x_m/x_0 , and the time ratio Δ_1 is fixed by

$$R(x_m) = \frac{1}{\sqrt{\mu}} \frac{\Delta_1^{1/2}}{1 - \Delta_1} \quad (20)$$

Together, these reduce the matching conditions to two algebraic equations for the two remaining unknowns α and Δ_2 . Their solution is $\alpha \approx \beta^{1/2} = 0.52 \times 10^{26}$ cm, and because later results are rather insensitive to small variations around this approximate equality we shall simplify by choosing $\alpha = \beta^{1/2}$, corresponding to $v = 1/2$. This fixes the adiabatic invariant $(\rho_\gamma/\rho_m)/T = \beta^{1/2}/2RT$, and hence the specific entropy per proton $S = 3.7 m_{pr} \rho_\gamma/\rho_m T = 3.7 \beta^{1/2} m_{pr}/2RT$. To find the present values of these two invariants we must know the present value of the conformal time x_0/x_s , which we can find from equation (11). With $\alpha = \beta^{1/2}$, this simplifies to

$$F(x) = \alpha(x_s^{-1} - x^{-1}) \quad (21)$$

so that the RW function becomes $R(x) = xF(x) = \alpha(x/x_s - 1)$. Our present location in cosmological history is determined by the value of the Hubble constant which, choosing $u = 1$, is $H_0 = 10^{-28}$ cm $^{-1}$. Expressed in conformal coordinates, the Hubble function is

$$H(x) = (\alpha x x_s)^{-1} (x_s^{-1} - x^{-1})^2 \quad (22)$$

so that at present $x_0/x_s \approx 10^2$. The present value for the RW function is therefore (see equation (16) with $\tilde{u} = 1$) $R_0 = 10^{28}$ cm. Finally, the determination of the cosmological invariant RT leads to a present value for the black body radiation background temperature of $T_0 = 11.8$ cm $^{-1} = 2.7$ K. We then have $\rho_\gamma/\rho_{m0} = 2.6 \times 10^{-3}$ and $S = 3.8 \times 10^{10}$. The consistency of the matching conditions and the ultra-relativistic nature of the RW medium at the matching point can be checked: from $(\rho_m/\rho_\gamma)_{\text{match}} = 2R(x_m)/\beta^{1/2}$ and $\rho_d = \rho_\gamma + \rho_m$, we have $\rho_\gamma = \rho_d(1 - 2R(x_m)/\beta^{1/2})^{1/2}$ or

$$T_\gamma = T_{bh} \left(1 - 2 \frac{R(x_m)}{\beta^{1/2}}\right)^{1/4} \approx T_{bh} \left(1 - \frac{R(x_m)}{2\beta^{1/2}}\right) \quad (23)$$

Because $R(x_m)/\beta^{1/2} = 10^{-27}$, this relation reflects a small ‘pollution’ of the radiation-dominated medium by massive particles, and also indicates near-continuity of the temperature at the matching point.

Conclusions

We have presented a singularity-free cosmological history marked by two successive phase transitions. The first is due to an instability of the vacuum and the second is the transition from an inflationary stage to the usual matter-radiation RW universe. The matching of the two regimes at the second transition shows that the de Sitter phase contains a dilute black hole gas characterized by the instability threshold mass M_{th} and whose evaporation time corresponds to the lifetime of the de Sitter phase. From the characteristics of the initial de Sitter stage we have deduced the value of the specific entropy per proton in the universe today. The primaeval stage has parameters depending only on the three fundamental constants h , c and κ , and the same is then true for the Universe today.

It is easy to extend our model to take into account, for example, the number N of helicity states of the primordial gas emitted by the black holes. The density at the matching point is then $\rho_d = (NC^4(N)/2)a(\kappa M_{th})^{-4}$, where before the quantity

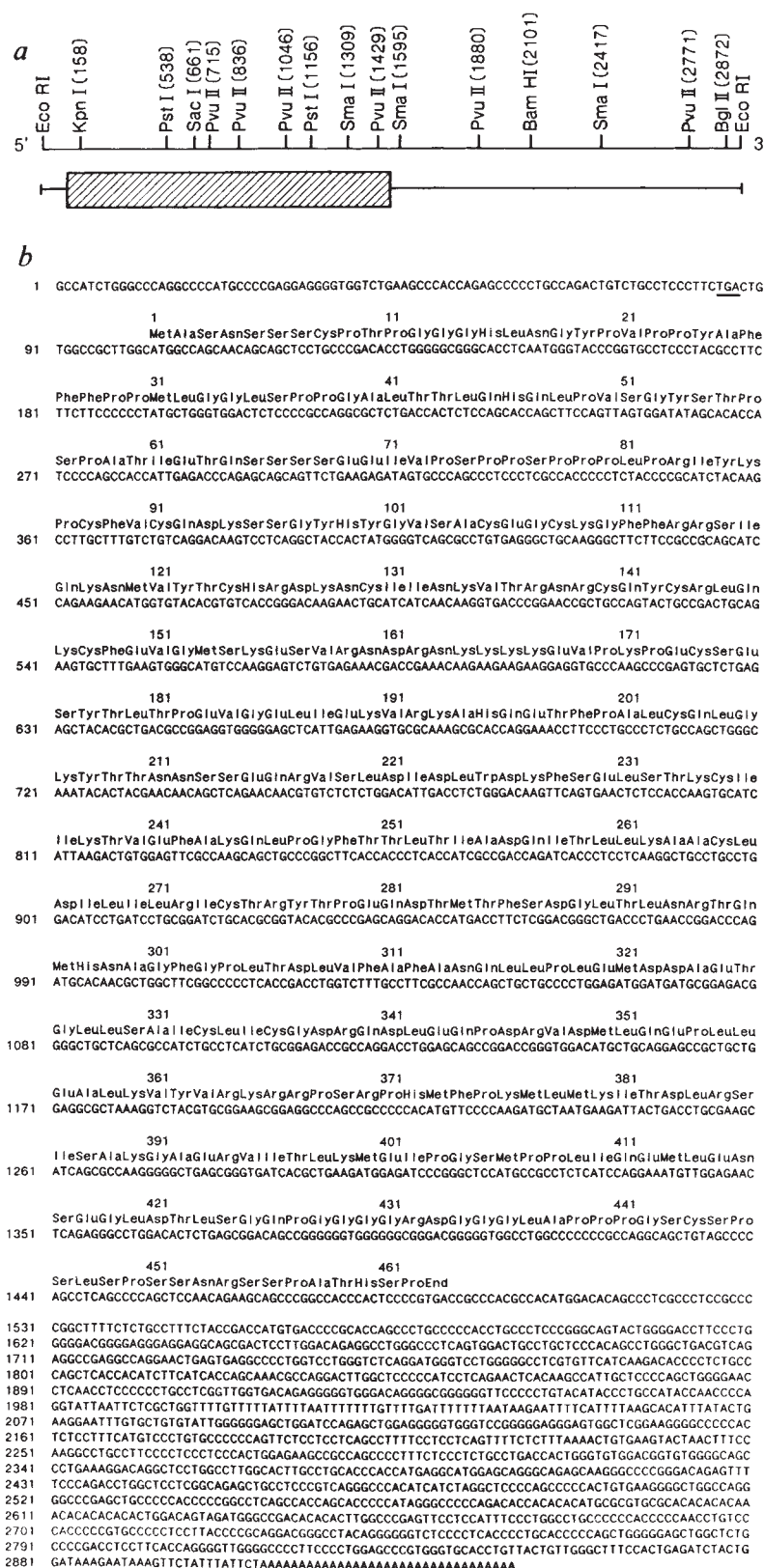


Fig. 1 DNA and primary amino-acid sequence of λ hK1R. *a*, Schematic representation and restriction enzyme map of the λ hK1R clone. The stippled box represents the predicted open reading frame. *b*, The complete nucleotide sequence of λ hK1R is shown with the predicted amino-acid sequence given above the long open reading frame. An upstream in-frame stop codon at nucleotides 85–87 and the putative polyadenylation signal are underlined.

Methods. A 63-mer oligonucleotide corresponding to nucleotides 408–477 of the genomic sequence published by Dejean *et al.*¹³ was used as a hybridization probe to screen a human testis λ gt10 library. The hybridization mixture contained 35% formamide, 1 $\times$ Denhardt's 5 $\times$ SSPE (0.15 M NaCl, 0.01 M Na₂HPO₄, 0.001 M EDTA), 0.1% sodium dodecyl sulphate (SDS), 100 μ g ml⁻¹ denatured salmon sperm DNA and 10⁶ c.p.m. ml⁻¹ of ³²P-labelled oligonucleotide. Duplicate nitrocellulose filters were hybridized at 42 °C for 16 h, washed three times for 20 min each in 2 $\times$ SSC, 0.1% SDS (1 $\times$ SSC is 150 mM NaCl, 15 mM sodium citrate) at 55 °C and autoradiographed at -70 °C with an intensifying screen. Clone λ hT1R obtained from this screening was partially characterized and then used as a hybridizing probe to screen a human kidney λ gt10 cDNA library³⁷. For this screening, the washing conditions were modified to 1 $\times$ SSC with 0.1% SDS at 68 °C. Several cDNA clones were isolated and the longest clone, λ hK1R, was digested with a number of restriction enzymes and the resulting fragments were subcloned in both orientations into the M13 sequencing vectors mp18 and mp19 and sequenced by the dideoxy procedure³⁸. DNA sequences were compiled and analysed by the programs of Devereux *et al.*³⁹ and Staden⁴⁰.

able⁵. Specifically, we have replaced the DNA-binding domain of the putative novel receptor with the well-described DNA-binding domain of the glucocorticoid receptor. This chimaeric construction, when expressed in cells, produces a hybrid receptor whose activation of a glucocorticoid-inducible promoter is dependent on the presence of the new ligand. Our studies indicate that this ligand is the vitamin A-related morphogen retinoic acid.

These data identify a presumptive specific high-affinity retinoic-acid receptor. The homology of this receptor to those of the steroid and thyroid hormones suggests a unifying hypothesis for both receptor structure and hormone action.

Candidate receptor

Analysis of the integration site of a hepatitis B virus from a human hepatocellular carcinoma led to the fortuitous iden-

Fig. 2 a, Construction of the chimaeric receptor hRGR. The domain-structure of the various constructions are shown schematically, the numbers correspond to the amino-acid positions of each domain. The DNA-binding domains are represented by 'DNA' and the ligand-binding domains by their respective inducers. The *NotI* and *XhoI* sites created by site-directed mutagenesis to permit the exchange of the DNA-binding domains between receptors are indicated. **b**, Induction of CAT activity by retinoic acid. The expression vectors were cotransfected into CV-1 cells with the reporter plasmid MTVCAT and cultured for 2 days in absence or presence of 100 nM dexamethasone (DEX) or retinoic acid (RA). The receptors inserted into the expression vectors are: pRShGR, human glucocorticoid receptor; pRShRR, human retinoic-acid receptor; pRShRR_{nx}, mutated human retinoic-acid receptor with *NotI* and *XhoI* sites; pRShRGR, chimaeric receptor composed of the human retinoic-acid receptor which DNA-binding domain has been replaced by the human glucocorticoid receptor DNA-binding domain.

Methods. **a**, Restriction enzyme fragments of the cDNA inserts of λ hK1R and hGR¹ were subcloned into the *KpnI* and *BamHI* sites of the mp19 vector and mutagenized according to the method of Kunkel⁴¹. The oligonucleotides used for the creation of the *NotI* site within hGR and hRR were 28 and 31 nucleotides respectively, whereas the oligonucleotides used for the creation of the *XhoI* site within hGR and hRR were 24 and 23 nucleotides. The creation of the *NotI* site resulted in the mutation of Pro 416 to an Arg residue in hGR_{nx}, and in the mutation of Ile 84 and Tyr 85 to Pro residues in hRR_{nx}. The introduction of the *XhoI* site did not alter the hGR_{nx} amino-acid sequence but resulted in the mutation of Lys 155 to a Leu residue in hRR_{nx}. The mutant receptors were then transferred to the expression vector pRS³, and the *NotI*-*XhoI* restriction fragment of pRShGR_{nx} containing the hGR DNA-binding domain was introduced into pRSRR_{nx} between the *NotI* and *XhoI* sites to create pRShRGR. **b**, Cell transfection and CAT assay. The recombinant DNA constructs (5 μ g each) were introduced into CV-1 cells by calcium phosphate coprecipitation⁴². The cells were cultured for two days in serum-free media supplemented with Nutridoma (Boehringer Mannheim) in presence or absence of inducers. CV-1 cells were then prepared for CAT assays as described⁴³ and the assays performed for 3 h using 25 μ g of protein extract. Experiments with retinol were conducted in subdued light.

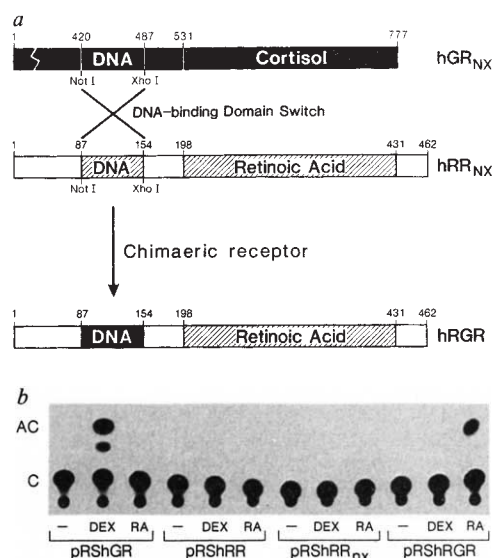
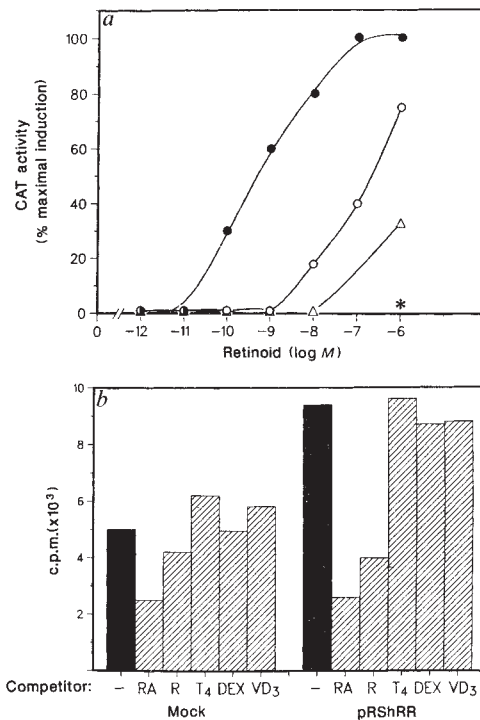


Fig. 3 a, Dose-response to retinoids. CV-1 cells cotransfected with pRShRGR and pMTVCAT were treated with increasing concentrations of retinoids or a single 1 μ M dose (*) of testosterone, dihydrotestosterone, oestrogen, cortisol, aldosterone, progesterone, triiodo-thyronine (T_3), thyroxine (T_4), dihydroxy-vitamin D_3 (VD_3) and 25-OH-cholesterol. The levels of CAT activity were plotted as percentages of the maximal response observed in this experiment. ●, Retinoic acid; ○, retinol; △, retinyl acetate or retinyl palmitate. **b**, Retinoic acid binding to cytosol extracts of transfected COS-1 cells. Bars represent bound ³H-retinoic acid determined in absence (black bars) or presence (stippled bars) of a 1,000-fold excess of various competitors. The values represent the mean of quadruplicate determinations. Competitors are retinoic acid (RA), retinol (R), T_4 , dexamethasone (DEX) and VD_3 .

Methods. **a**, CV-1 cell cotransfections and CAT assays were performed as described in Fig. 2. Retinoic acid was dissolved in a minimum volume of dimethyl sulphoxide and diluted in ethanol. All other products were diluted in ethanol and control cultures received 0.1% solvent (v/v) in media. Dose-response curves were performed in triplicate. **b**, Subconfluent COS-1 cells were transfected with 10 μ g per dish of a control plasmid (pRS) or pRShRR by the DEAE-Dextran method⁴⁴. Cells were maintained for 2 days in Dulbecco's minimal Eagle's medium (DMEM) with 5% charcoal-treated fetal calf serum, then harvested in TNE (40 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA) and lysed by Dounce homogenization in hypotonic buffer (50 mM Tris-HCl pH 7.4, 0.1 mM EDTA, 5 mM dithiothreitol, 10 mM NaMoO₄, 10% glycerol, 0.5 mM phenylmethylsulphonyl fluoride) and centrifuged at 100,000g for 30 min to yield the cytosol fraction. Incubations were performed in hypotonic buffer with 150 μ g of protein from the cytosolic fraction and 2×10^{-8} M ³H-retinoic acid (NEN, 52.5 Ci per mmole) in a total volume of 200 μ l. Specific binding was measured by the addition of 2×10^{-5} M of competitors. Reactions were carried out at 4°C for 16 h. Bound ³H-retinoic acid was measured using DE-81 filters. Reactions were placed on filters for 1 min, rinsed with 5 ml of washing buffer (50 mM Tris-HCl pH 7.4, 0.1 mM EDTA, 0.1% Triton X-100), dried and counted by liquid scintillation spectrophotometry.



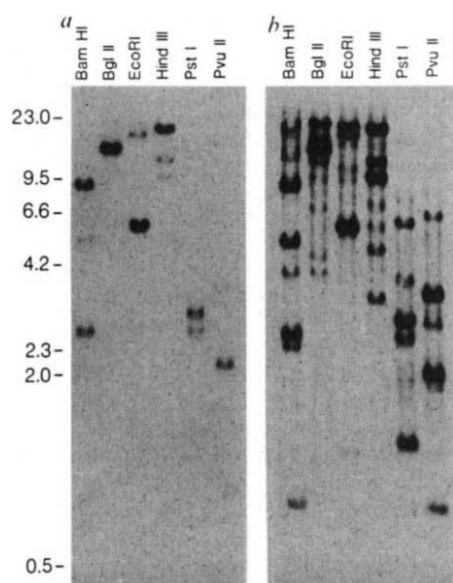


Fig. 4 Southern blot analysis of human genomic DNA. *a*, Human placenta DNA was digested with the indicated restriction enzymes. After separation of the digested DNA in a 0.8% agarose gel (10 μ g per lane) and transfer to nitrocellulose filters⁴⁵, the blots were hybridized with an *EcoRI*-*PvuII* fragment from λ hT1R (~600 bp) encompassing the DNA-binding domain of the hRR under high stringency conditions (50% formamide, 5 $\times$ SSPE, 1 $\times$ Denhardt's, 0.1% SDS, 100 μ g ml⁻¹ salmon sperm DNA). The filter was washed in 0.1 $\times$ SSC, 0.1% SDS at 65 $^{\circ}$ C. The λ *HindIII* DNA markers (size in kb) are aligned to left of the autoradiograph. *b*, Analysis of human placenta DNA using the same probe as in *a* under non-stringent conditions. A parallel blot containing identical samples was hybridized as in *a*, except that 35% formamide was used. The filter was washed in 2 $\times$ SSC, 0.1% SDS at 55 $^{\circ}$ C.

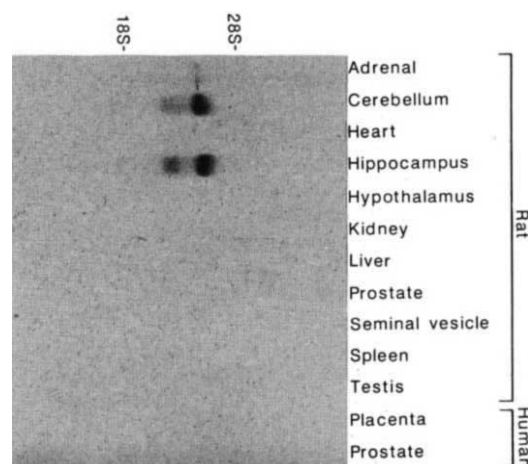


Fig. 5 Northern blot analysis of retinoic acid receptor mRNA in rat and human tissues.

Methods. Total RNA was isolated from various tissues using guanidine thiocyanate⁴⁶, 20 μ g were separated on a 1% agarose-formaldehyde gel, transferred to a nitrocellulose filter, and hybridized under stringent conditions using the probe described in Fig. 4. Ribosomal RNAs (28S and 18S) are indicated for size markers. The nitrocellulose filter was autoradiographed at -70 $^{\circ}$ C with an intensifying screen for 1 week.

tification of a novel genomic sequence with striking similarity to the DNA-binding domain of the steroid hormone receptors¹³. To examine the possibility that this gene encodes a previously unknown receptor, an oligonucleotide derived from this sequence was labelled and used to probe a number of human cDNA libraries. Five positive clones were initially isolated from a testis cDNA library. The insert from one of these clones (λ hT1R) was used to isolate additional cDNA clones from a λ gt10 kidney cDNA library. A restriction map of the largest clone (λ hK1R) is shown in Fig. 1*a*. Nucleotide sequence analysis reveals a long open reading frame of 462 amino acids beginning with a presumptive initiator methionine codon corresponding to nucleotides 103-105 (Fig. 1*b*). The sequence surrounding this ATG agrees with the consensus described by Kozak¹⁴ for a translation initiation site. Upstream of the ATG is an in-frame terminator providing support for the initiator methionine. Another methionine found 30 codons downstream does not conform to the consensus and is an unlikely initiator. After the terminator codon at position 1,489-1,491 is a 1,419 nucleotide 3'-untranslated region with a consensus polyadenylation signal (AATAAA) found 20 nucleotides upstream of a polyadenylated tract¹⁵.

A polypeptide of relative molecular mass 50,772 (M_r 51K), is encoded within the translational open reading frame. The size of the protein encoded by the insert of λ hK1R was verified by *in vitro* translation of RNA¹⁶ derived from this insert and found to correspond to the predicted size of 54K (data not shown). Amino acid sequence of this protein has been compared to the glucocorticoid and thyroid hormone receptors. The highest degree of similarity is found in a cysteine-rich sequence of 66 amino acids beginning at residue 88. We have previously shown that this region of the hGR is the DNA-binding domain^{3,4}. In

addition, mutagenesis and expression studies have provided direct evidence for its role in transcriptional activation of genes harbouring glucocorticoid response elements (GREs)^{3,4}.

Ligand identity

As the ligand for the gene product of λ hK1R was unknown, we wished to develop a quick and sensitive assay to reveal its identity. The DNA-binding domains of the human glucocorticoid and oestrogen receptors can be interchanged, yielding functional hybrid receptors. One such chimera recognizes the glucocorticoid-responsive element of the MMTV-LTR but stimulates transcription in an oestrogen-dependent fashion⁵. This suggests a general strategy that can be exploited to identify the ligand-binding properties of a novel hormone receptor. To test this approach we have substituted the DNA-binding domain of the λ hK1R gene product with the DNA-binding domain from the hGR (Fig. 2*a*). The assay system is established by transfecting CV-1 cells with the hybrid receptor gene and a MMTV-CAT reporter gene. Transfected cells are then systematically challenged with a battery of candidate ligands and induction monitored by changes in chloramphenicol acetyltransferase (CAT) activity. Because of their hormone-like activities, the retinoids, including retinol (vitamin A) and retinoic acid, were evaluated as potential inducers. Remarkably, retinoic acid elicited a dramatic increase in CAT activity of the hybrid receptor (Fig. 2*b*). No effect upon CAT activity was observed using the parent vector, pRShRR_{nx}, or the wild-type gene product from λ hK1R, here referred to as the human retinoic acid receptor (hRR). As expected, the hybrid receptor is not induced by glucocorticoids, and the hGR is not induced by retinoic acid.

As shown in Fig. 3*a*, retinoic acid gives an ED₅₀ value of 6×10^{-10} M on CAT activity induced by the hybrid receptor, which is consistent with ED₅₀ values observed for retinoic acid in a variety of biological assays¹⁷. Retinol functions as a weak agonist with an ED₅₀ value greater than 100 nM. Retinyl acetate and retinyl palmitate are even weaker inducers. A number of natural and synthetic ligands including testosterone, dihydrotestosterone, oestrogen, dexamethasone, cortisol, aldosterone, progesterone, T₃, T₄, Vitamin D₃ and 25-OH-cholesterol did not induce CAT activity.

Our hybrid receptor activation assay allows us to screen a large number of potential activators, but does involve the use

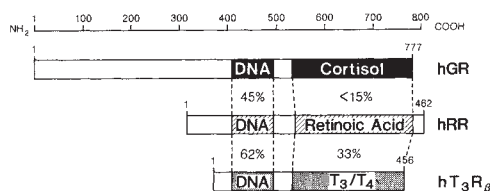


Fig. 6 Schematic amino-acid comparisons of the hGR, hRR and hT₃R_β structures. Amino-acid sequences have been aligned schematically within the percentage amino-acid identity for each region of similarity in the intervals between dotted lines.

of a hybrid protein. To corroborate the identity of the λ hK1R gene product as the retinoic acid receptor, the binding properties of the expressed protein were evaluated following transfection of COS-1 cells. As shown in Fig. 3b, transfected cells reveal increased capacity to bind [³H]retinoic acid. This increase occurs over an endogenous background that is probably a consequence of the presence of cellular retinoid-binding proteins (see discussion) as well as nonspecific binding. Consistent with the activation studies, the binding can be completely inhibited by competition with unlabelled retinoic acid but only partially by retinol. Thyroid hormones, dexamethasone and vitamin D₃ did not compete with the binding of retinoic acid.

A gene family

To determine whether this gene is unique and to identify potentially related genes, human DNA was examined by Southern blot analysis. Hybridization of restriction endonuclease-digested human DNA with a labelled DNA fragment derived from the coding region of the hRR polypeptide produced three bands in every digestion consistent with a single hybridizing genetic locus (Fig. 4a). This hybridization pattern is unrelated to the restriction endonuclease map described by Dejean *et al.* for the hepatitis B virus pre-integration site¹³. When the hybridization conditions were relaxed, however, additional bands were observed in the products of each enzyme digestion (Fig. 4b). These observations indicate that there is another locus, and possibly more, in the human genome related to the retinoic-acid receptor.

Expression of the hRR gene

Because retinoic acid is known to exert effects on a large number of different cell types, we examined the expression of the hRR gene. Total cytoplasmic RNAs isolated from a variety of rat and human tissues were size-fractionated and transferred to a nitrocellulose filter. Hybridization with a 600-base pair (bp) restriction fragment from λ hT1R reveals a major RNA species of 3,200 nucleotides with highest levels in the hippocampus, adrenals, cerebellum, hypothalamus and testis (Fig. 5). Longer exposure shows that most tissues contain a small amount of the 3.2-kilobase (kb) transcript, although it is undetectable in some tissues such as liver. The size of the messenger RNA indicates that we have isolated a nearly full-length hRR cDNA.

Conclusions

The data in this report identify the gene product of λ hK1R as the human retinoic-acid receptor based on three criteria. First, the overall structural similarity of the hRR to steroid and thyroid hormone receptors (Fig. 6) suggests that it is likely to be a ligand-responsive regulatory protein. Second, an expressed chimaeric receptor, consisting of the hGR DNA-binding domain and the presumptive ligand-binding domain of the hRR, can act as a transcriptional regulator of a glucocorticoid-inducible reporter gene only in response to retinoic acid. This induction occurs at physiological levels. Third, expression of the candidate hRR in transfected cells selectively increases the capacity of those cells to bind retinoic acid.

Development and oncogenesis. The retinoids comprise a group of compounds including retinoic acid, retinol (vitamin A) and a series of natural and synthetic derivatives that together exert profound effects on development and differentiation in a wide variety of systems^{18–22}. Although early studies focused on the effects of retinoids on epithelial growth and differentiation, their actions have been shown to be more widespread than previously suspected. Many recent studies have examined the effects of these molecules on a variety of cultured cell lines including neuroblastomas²³, melanomas²⁴ and fibroblasts²⁵. In the human promyelocytic leukaemia cell line, HL-60, retinoic acid is a potent inducer of granulocyte differentiation²⁶. In F9 teracarcinoma stem cells, retinoic acid will induce the differentiation of parietal endoderm, characteristic of a late mouse blastocyst^{27–29}. Retinoic acid has been shown to exert equally potent effects in development. For example, in the developing chick limb bud, retinoic acid substitutes for the action of the polarizing region in establishing the anterior-posterior axis³⁰. By controlling exposure to retinoic acid, it is possible to generate novel patterns of limb structure. Although retinoic acid is primarily considered a morphogen, Northern blot analysis suggests a re-evaluation of its function in the adult. In humans, retinol deficiency has been linked to an alarming increase in a variety of cancers³¹. Retinoids have also been shown to inhibit tumour progression in animals and block the action of tumour promoters *in vitro*. In this context, the hRR may be considered as a negative regulator of oncogenesis.

The identification of a cellular retinol-binding protein (CRBP) and a distinct cellular retinoic-acid binding protein (CRABP) in the mid-1970s led to a proposal that they might represent specific intracellular receptor systems (see ref. 32 for review). These molecules are relatively small (134 amino acids for CRBP) and bear no obvious structural homology to the steroid receptors. More importantly, they show no similarity to the retinoic-acid receptor identified in this manuscript. Despite a detailed biochemical characterization and their recent cloning^{32–36}, there is no direct evidence that establishes a decisive role for CRBP and CRABP as mediators of retinoid action.

A superfamily of regulatory genes. Two surprising results have emerged from the studies presented here. The first is the discovery of a family of retinoic-acid receptor-related genes, implying the existence of one or more other proteins with closely related properties. Physiological studies demonstrate that both retinoic acid and retinol (vitamin A) can exert potent effects on cellular differentiation and that these effects are often not linked. It thus seems likely that at least one related gene product might be a specific retinol receptor or a receptor for another member of the retinoid family. The second surprising observation from these results is the close kinship of the retinoid receptor with the thyroid hormone receptor. This relationship is surprising in part because of the structural dissimilarity of the thyroid hormones and the retinoids. Thyroid hormones are derived from the condensation of two tyrosine residues whereas the retinoids are derived from mevalonic acid. The observation that chemically distinct molecules interact with receptors sharing common structures probably reflects a common mode of action with which they elicit their particular regulatory effects. Based on this analogy, we can now propose that the interaction of retinoids with their intracellular receptors induces a cascade of regulatory events that results from the activation of specific sets of genes by the hormone/receptor complex. Although animals employ diverse means to control their development and physiology, the demonstration that the retinoic-acid receptor is part of the steroid receptor superfamily suggests mechanisms controlling morphogenesis and homeostasis may be more universal than previously suspected.

We thank Jeff Arriza for the plasmid RShGR_{nx} and Dr Kazuhiko Umesono for expert advice and assistance in sequencing intact plasmid DNA. We also thank Dr Cory Weinberger and the members of the Evans laboratory for discussion, Dr

Donald Gruol for advice on hormone binding and Drs Bart Sefton, Tony Hunter and Michael McKeown for suggestions on the manuscript. We also thank Elaine Stevens for expert administrative assistance and help in the preparation of the manuscript. V.G. is a Centennial Fellow of the Medical Research

Council of Canada. R.M.E. is an investigator of the Howard Hughes Medical Institute. This work was supported by the Howard Hughes Medical Institute and by grants from the National Institutes of Health.

Received 13 November; accepted 24 November 1987.

- Hollenberg, S. M. *et al. Nature* **318**, 635-641 (1985).
- Green, S. *et al. Nature* **320**, 134-139 (1986).
- Giguere, V., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. *Cell* **46**, 645-652 (1986).
- Hollenberg, S. M., Giguere, V., Segui, P. & Evans, R. M. *Cell* **49**, 39 (1987).
- Green, S. & Chambon, P. *Nature* **375**, 75-78 (1987).
- Kumar, V., Green, S., Stark, G., Berry, M., Jin, J.-R. & Chambon, P. *Cell* **51** (in the press).
- Miesfeld, R., Godowski, P. J., Maler, B. & Yamamoto, K. R. *Science* **236**, 434-427 (1987).
- Weinberger, C., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. *Nature* **318**, 670-672 (1985).
- Weinberger, C. *et al. Nature* **324**, 641-646 (1986).
- Sap, J. *et al. Nature* **324**, 635-640 (1986).
- Arriza, J. *et al. Science* **237**, 268-275 (1987).
- Thompson, C. C., Weinberger, C., Lebo, R. & Evans, R. M. *Science* **237**, 1610-1614 (1987).
- Dejean, A., Bougueret, L., Grzeschik, K.-H. & Tiollais, P. *Nature* **322**, 70-72 (1986).
- Kozak, M. *Nucleic Acid Res.* **16**, 8125-8148 (1987).
- Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211-214 (1976).
- Krieg, P. A. & Melton, D. A. *Nucleic Acid Res.* **12**, 7057-7070 (1984).
- Sporn, M. B. & Roberts, A. B. in *The Retinoids* Vol. 1 (eds Sporn, M. B., Roberts, A. B. & Goodman, D. S.) 235-279 (Academic, New York, 1984).
- Sporn, M. & Roberts, A. B. *Cancer Res.* **43**, 3034-3040 (1983).
- Mandel, G. & Cohen, V. in *The Pharmacological Basis of Therapeutics* (eds Gilman, A., Goodman, L., Rall, T. & Mural, F.) 1573-1591 (Macmillan, New York, 1985).
- Wolback, S. B. & Howe, P. R. *J. exp. Med.* **62**, 753-777 (1925).
- Lotan, R. *Biochim. biophys. Acta* **605**, 23-91 (1980).
- Fuchs, E. & Green, H. *Cell* **25**, 617-625 (1980).
- Haussler, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 5525-5529 (1983).
- Lotan, R., Stolarsky, T. & Lotan, D. *Cancer Res.* **43**, 2868-2875 (1983).
- Shroder, E., Rapaport, E., Kabeenell, K. & Black, P. H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1549-1552 (1982).
- Breitman, T., Selonick, S. & Collins, S. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2936-2940 (1980).
- Strickland, S. & Mahdavi, V. *Cell* **15**, 393-403 (1978).
- Jetten, A., Jetten, M. & Sherman, M. *Exp. Cell Res.* **124**, 381-392 (1979).
- Wang, S.-Y., LaRosa, G. & Gudas, L. J. *Dev. Biol.* **107**, 75-86 (1985).
- Thaller, C. & Eichele, G. *Nature* **327**, 625-628 (1987).
- Moon, R. C. & Itri, L. M. in *The Retinoids* (eds Sporn, M. B., Roberts, A. B., & Goodman, D. S.) 327-371 (Academic, New York, 1984).
- Chytil, F. & Ong, D. E. in *The Retinoids* Vol. 2 (eds Sporn, M. B., Roberts, A. B., Goodman, D. S.) 89-123 (Academic, New York, 1984).
- Sundelin, J., Das, S. R., Eriksson, U., Rask, L. & Peterson, P. A. *J. biol. Chem.* **260**, 6494-6499 (1985).
- Saari, J., Futterman, S. & Bredberg, L. *J. biol. Chem.* **253**, 6432-6436 (1978).
- Sundelin, J. *et al. J. biol. Chem.* **260**, 6488-6493 (1985).
- Colantuoni, V. *et al. Biochem. biophys. Res. Commun.* **130**, 431-439 (1985).
- Bell, G. I. *et al. Nucleic Acid Res.* **14**, 8427-8446 (1986).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Devereux, J., Haebler, P. & Smithies, O. *Nucleic Acid Res.* **12**, 387-395 (1984).
- Staden, R. *Nucleic Acid Res.* **10**, 2951-2961 (1982).
- Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488-492 (1985).
- Wigler, M. *et al. Cell* **16**, 777-785 (1979).
- Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
- Deans, R. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 1292-1296 (1984).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Chirgwin, J. M., Przybyla, A. F., McDonald, R. J. & Rutter, W. F. *Biochemistry* **18**, 5294-5299 (1980).

LETTERS TO NATURE

How accurate were seventeenth-century measurements of solar diameter?

C. R. O'Dell & A. Van Helden

Scientia, Rice University, PO Box 1892, Houston, Texas 77251, USA

The assumption that the solar diameter is constant over periods that are short compared to the nuclear process has recently been questioned. In 1979 Eddy and Boornazian claimed that measurements of the solar diameter made at the Greenwich and US Naval observatories show a decrease of ~ 2 arc s, or 0.1%, per century over the period 1836-1953 (refs 1, 2). Others have since extended this baseline by studying the evidence from Mercury transits³ and assessing the entire range of evidence from the beginning of the eighteenth century to the present⁴, and have found little or no evidence for a secular change in the solar diameter. Most recently, Ribes *et al.*⁵ have pushed the time baseline back to 1666, by considering the micrometer and transit measurements of the solar diameter made in Paris by Jean Picard (1666-82) and Philippe and Gabriel-Philippe de La Hire (1683-1719). Not only are these observations of importance for the investigation of solar changes at the end of the Maunder minimum, they can also be used for comparison with modern values. We discuss here the accuracy of Jean Picard's micrometer measurements of the solar diameter and his successors' transit measurements in light of the claim by Ribes *et al.*⁵ that these data provide evidence for a change in the apparent size of the Sun at the end of the Maunder minimum and a difference from the current size. Our evidence suggests that the necessary corrections to the measurement of sizes with early telescopes are larger than Ribes *et al.* assume. Therefore we call into question their conclusion that the Sun rapidly changed in size during this period.

Ribes *et al.*⁵ argue that their analysis of the measurements of Picard and Philippe and Gabriel-Philippe de la Hire shows that

the observed solar diameter, uncorrected for systematic errors, was $32' 9''$ for the period 1666-82, and then decreased smoothly to $32' 6''$ over the period 1683-1718. The present-day value is $31' 59.3''$ (ref. 6). The 1666-82 period corresponds to the deep Maunder minimum and the 1683-1718 period to the end of the Maunder minimum; micrometer measurements make up the bulk of the data from the former, while transit measurements constitute the data from the latter period. While the accuracy of recent observations can be assessed with some precision, the accuracy of older observations presents severe problems.

Jean Picard (1620-82) was a pioneer in the precision astronomy made possible by the micrometer and the application of telescopic sights to measuring arcs. Picard's measurements of the diameters of the Sun, Moon and planets, begun in 1666, were the first sustained series of micrometer measurements made. They survive in manuscript form in the Paris Observatory and were published in 1741 in P-C Le Monnier's *Histoire Céleste*⁷. In assessing the accuracy of these measurements, two aspects must be considered, the quality of the image and the accuracy of the micrometer.

The quality of the image will significantly affect the measurement of solar diameter, regardless of whether the image was projected or viewed directly through dark glass. The image formed is a product of the observing environment and the telescope itself. Each component will produce an essentially circular disk (ignoring lens astigmatism) with a peak surface brightness at the middle and a monotonically decreasing surface brightness with an approximately gaussian distribution. There are several ways of characterizing such an image, the most convenient being the full width at half the maximum intensity (FWHM). If the image is made up of more than one component, each component will add quadratically to the total; for example, two equal components would increase the resultant image size by a factor of 1.414. The observing environment produces the phenomenon of astronomical seeing, occasioned by inhomogeneities in the refractive index of the air along the line of sight, caused by small variations in temperature. Very small

aperture telescopes, such as were used in the seventeenth century, manifest seeing primarily by random variations in position, but intense heating conditions will blur the image. The result is that for the best night-time observations the seeing effect will be unimportant, the image quality being determined primarily by the telescope itself, while daytime observations will have a seeing component of the FWHM which will be ~ 3 arc s in size.

As Picard's anachromatic telescope used simple biconvex lenses, computation of image quality is straightforward. Unfortunately, however, we know little about the instrument or instruments Picard used, except that, as Ribes *et al.* report⁵, the focal length of his transit telescope was 6 feet. If the focal ratio (f) was ~ 30 , typical for telescopes of that era, the aperture would have been ~ 2.4 inches, from which we can compute the image properties. These calculations, for images on the axis as well as at the solar radius, were made by D. J. Schroeder, using his optical programs applied to the design of the Hubble Space Telescope. The spherical aberration broadens the image by ~ 0.8 arc s with only a minor wavelength dependence. The more important aberration is chromatic, which is very serious: it alone broadens the image by 72 arc s for each 100-nm wavelength interval at $f30$ in the case of typical crown glass. As the effective bandwidth of the eye-telescope combination would have been ~ 200 nm, image quality must have been very poor. The observer was thus forced to select a colour for focusing the image, while letting other colour images become totally blurred. The very best image would have been limited by the smallest wavelength range discernible by the eye and must have been quite large. Moreover, these modern calculations of the image quality assume that the lens curvatures were perfectly spherical and that the optical quality of the glass itself was optimal. Both assumptions must be treated with extreme caution, for surviving examples of seventeenth-century lenses often have irregularities in curvatures and imperfections in the glass. In view of these considerations, therefore, it is too optimistic to say, as Ribes *et al.* do⁵, that the error introduced by spherical and chromatic aberration is "unlikely to have been more than one arcsecond". Ribes *et al.* are certainly aware of such optical effects and describe them with the term irradiation. They assume that the irradiation correction to the solar diameter is 3.1 arc s, as derived from late nineteenth-century telescopes⁸. We question the relevance of applying corrections derived from telescopes designed and built two centuries after the Picard instruments. The objective lenses of late nineteenth-century telescopes would have been colour-corrected doublets whose image quality was limited only by astronomical seeing.

It is difficult quantitatively to assess the resulting effective FWHM. We note, however, that only once in his observing record, in 1673, does Picard report (ref. 7, p. 28) that he "... began to perceive that the disk of Jupiter was a bit oval and that the greatest diameter is always along the bands". (The oblateness of Jupiter's disk had been discovered by Picard's colleague, Giovanni Domenico Cassini, ~ 1665 .) Near opposition Jupiter's disk is non-circular by ~ 3 arc s, and we can thus infer that the night-time image was ~ 3 arc s, and that the daytime image was even worse, perhaps 5 arc s. Clearly, then the instrumental image seriously affected the apparent size ascribed to an astronomical object viewed through such a telescope as Picard's. In the case of viewing stars, whose intrinsic size is much smaller than the instrumental image, one would ascribe an apparent diameter equal to the instrumental image (although most seventeenth-century observers were aware that this was a spurious disk). For an object subtending a resolved disk, the ascribed apparent diameter would be the true disk diameter plus a fraction of the FWHM of the instrumental image. For a faint disk, the overestimate would be close to the value of the FWHM, and for a bright disk the image would be filled in more and could be overestimated by a value greater than the FWHM, as the eye would then detect the boundary further out on the image profile.

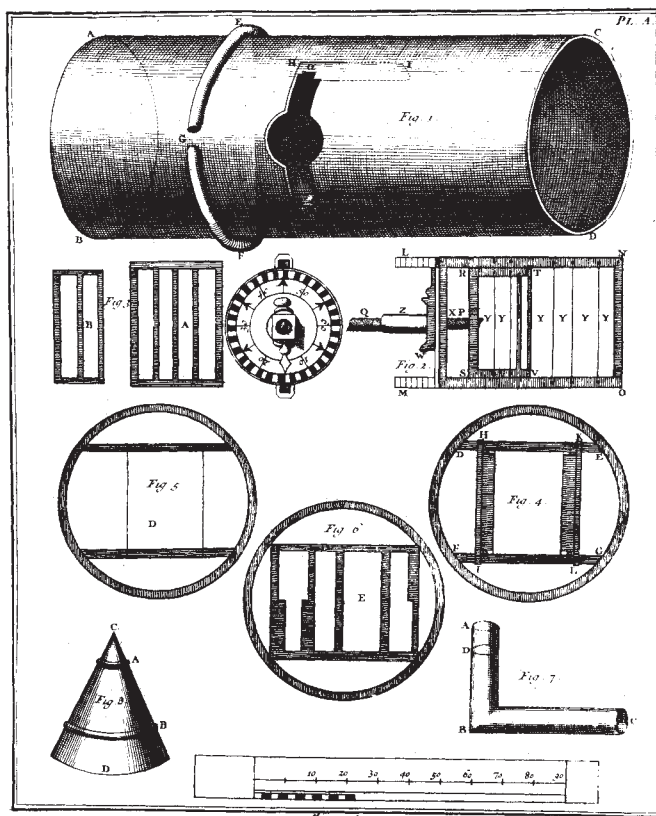


Fig. 1 Illustration of the micrometer used by Jean Picard for making the first accurate observations of the solar diameter, as depicted by his collaborator Adrien Auzout¹⁰. Although this instrument represented a vast improvement in observational capability, significant random and systematic errors in the results obtained were still a problem.

The accuracy of the micrometer used by Picard is even more difficult to assess. Ribes *et al.* ascribe a "potential accuracy" of 1 arc s to this instrument. Picard worked together with his colleague Adrien Auzout in the development of a filar micrometer in 1666. Auzout discussed this instrument at a meeting of the Académie Royale des Sciences in 1667 (ref. 9) and published a memoir on it a few months later¹⁰. Figure 1 shows this instrument, and its workings. Errors in this instrument could result from inconsistencies in the screw thread as well as play between the various parts. This may be why Picard initially took the micrometer frame out of the telescope to measure the distance between the hairs¹¹.

Astronomers were aware that, although this new instrument gave greater precision, it was not perfect. Even in his most enthusiastic utterings, Auzout did not claim the accuracy that Ribes *et al.* assign to this instrument. On 28 December 1666 Auzout wrote¹²:

I did apply my self the last Summer to the taking of the *Diameters* of the Sun, Moon, and the other Planets, by a Method, which one M. Picard and my self have, esteem'd by Us the best of all those, that have been practis'd hitherto; since we can take the *Diameters* to *Second Minutes* [that is, arcseconds], being able to divide one foot into 24000. or 30000. parts, scarce failing as much as in one only part, so as we can in a manner be *assur'd*, not to deceive our selves in 3. or 4. seconds.

Auzout's estimate of the accuracy of micrometer measurements thus confirms our suspicions. Can we test these assessments of the correction factor by examining actual

measurements of an object of known angular size? Fortunately there is an appropriate set of observations. As well as the solar measurements cited by Ribes *et al.*⁵, Picard's observation record contains micrometer measurements of the Moon and all the planets then known. Of these, his measurements of Jupiter's angular diameter present the best test case: Jupiter presents a comparatively large disk, its surface brightness is not great, and its angular diameter on a given day can be precisely computed. Picard measured this diameter on 18 occasions between 3 October 1666 and 10 December 1676 (gregorian dates), a period that partly coincides with the interval of solar observations that forms the core of the micrometer observations in the discussion by Ribes *et al.* The reasonable supposition that he measured the polar diameter by aligning the micrometer hairs along the equator, allowing the planet's image to move through the field and graze the micrometer threads, is supported by Auzout's remarks¹³.

Picard's measurements are compared with calculated values in Table 1. We have used a detailed ephemeris of Jupiter calcu-

the difference from the modern value of 959.63 arc s (ref. 6) to between only 0.8 to 2.8 arc s, much less than argued by Ribes *et al.*, and certainly within the remaining systematic errors in the micrometer screw. A difference of 1.8 arc s corresponds to a screw error of 1 part in 533, which is certainly possible. Although seventeenth-century instrument makers were supremely skilled, description of screw-cutting methods then in use indicate that micrometer screws probably suffered from significant errors¹⁴. Such errors would be relatively insignificant over Jupiter's small angular diameter, but would be much larger over that of the Sun.

Ribes *et al.* also assert that transit observations made over the interval marking the end of the Maunder minimum provide evidence that the Sun rapidly contracted at this time. As the observations during the deep Maunder minimum were largely made with the micrometers, it seems an uncertain assertion to compare the results of two very different techniques. Transit observations are subject to very significant observer biases arising from systematic judgement errors in measuring an event. Moreover, the transit observations will be subject to the same overestimate of the image size due to the image quality. We find an average observed solar radius value of 963.54 ± 0.12 arc s for the period 1684–1719 from the annual uncorrected averages given by Ribes *et al.* in their Table 1. There is no statistically significant variation in the interval 1684–1719. If the night-time correction of 2.1 ± 0.3 arc s determined from the jovian observations is applied, then the transit values are reduced to 961.4 ± 0.3 arc s, which is to be compared to the modern value of 959.63 arc s. The disagreement of 1.8 arc s is probably due to the fact that day-time image correction would be larger than the night-time value of 2.1 arc s, and there is probably an additional and unknown correction for observer bias. The few micrometer measurements made from 1710 to 1713 show a large uncertainty and do not provide an independent confirmation. Our conclusion from the correction of the transit observations is that they give results in agreement with the modern values. This position is strengthened by the very sensitive 1715 solar eclipse observation of Halley^{4,15}, which leads to an estimated solar radius of 960 arc s, the same as the modern value.

We conclude that the early measurements of the solar diameter by Jean Picard and his immediate successors must be corrected for systematic instrumental effects whose optical component can be estimated from similar observations made of Jupiter with the same or similar instruments. The small remaining difference between the corrected micrometer measurements during the deep Maunder minimum and the modern values can be attributed to errors in the micrometer. The optically corrected transit observations are in agreement with the modern values.

We thank D. J. Schroeder of Beloit College and D. K. Yeomans of the Jet Propulsion Laboratory for their calculations, and R. L. Gilliland of the National Center for Atmospheric Research and L. V. Morrison of the Royal Greenwich Observatory for their comments.

Received 22 June; accepted 19 October 1987.

1. Eddy, J. A. & Boornazian, A. A. *Bull. Am. astr. Soc.* **11**, 437 (1979); *Phys. Today* **32**(9), 17–19 (1979).
2. Gilliland, R. L. *Astrophys. J.* **248**, 1144–1155 (1981).
3. Shapiro, I. I. *Science* **208**, 51 (1980).
4. Parkinson, J. H., Morrison, L. V. & Stephenson, F. R. *Nature* **288**, 548–551 (1980).
5. Ribes, E., Ribes, J. C. & Bartholot, R. *Nature* **326**, 52–55 (1987).
6. Allen, C. W. *Astrophysical Quantities* 3rd edn, 162 (Athlone, London, 1973).
7. Le Monnier, P. C. *Histoire Céleste* (Paris, 1741).
8. Auwers, A. *Asiro. Nachr.* **128**, 361–365 (1891).
9. Académie des Sciences, Paris *Procès-Verbaux*, **2**, 50–58 (1667); printed in McKeon, R. M. *Physica* **13**, 247–251 (1971).
10. Auzout, A. *Traité du Micromètre* (Paris, 1667); reprinted in *Mémoires de l'Académie Royale des Sciences depuis 1666 jusqu'à 1699* vol. 7, part 1 (Paris, 1729).
11. McKeon, R. M. *Physica* **13**, 250 (1971).
12. Auzout, A. *Phil. Trans. R. Soc.* **21**, 373 (1667).
13. McKeon, R. M. *Physica* **13**, 249–251 (1971).
14. Daumas, M. *Scientific Instruments of the Seventeenth and Eighteenth Centuries* (translated by Holbrook, M.) 113–114 (Praeger, New York, 1972).
15. Dunham, J. B., Dunham, D. W. & Fiala, A. D. *Bull. Am. astr. Soc.* **12**, 832 (1980).

Table 1 Picard's measurements of Jupiter's angular diameter

Gregorian date	Angular diameter measured by Picard (arc s)	Calculated polar angular diameter (arc s)	Geocentric distance (AU)
1666 Oct. 3	51	46.23	4.0027
15	50	45.32	4.0840
26	50	44.17	4.1906
Nov. 10	45	42.30	4.3760
26	41	40.16	4.6089
Dec. 6	40	38.84	4.7649
9	40	38.46	4.8124
22	39	36.89	5.0179
1667 Feb. 20	36	32.02	5.7793
Sept. 11	50	43.80	4.2257
Oct. 13	54	46.36	3.9925
Nov. 1	54	46.37	3.9918
10	54	45.93	4.0301
1668 Oct. 19	52	43.10	4.2934
1673 Apr. 1	44	41.60	4.4495
13	44	41.98	4.4620
June 24	39	35.30	5.2441
1676 Sept. 10	42	39.90	4.6282

Taken from ref. 7. We have assumed a semidiameter at 1.00 AU of 98.37 arc s and an ellipticity of 0.063 (ref. 6, p. 140).

lated for us by D. K. Yeomans to compare the measured and true angular sizes (ref. 6, p. 140). We find that the observed diameter was larger by 4.27 ± 0.61 arc s, which is in good agreement with our expectation from night-time instrumental image size. The standard deviation of these measurements is 2.57 arc s, which probably represents variations in the seeing disk, observer errors and micrometer errors.

Picard's observations of the solar diameter must be corrected for the instrumental image. We base our corrections on the contemporaneous observations of Jupiter, while Ribes *et al.* use corrections determined for telescopes made two centuries later⁸. The Jupiter observations indicate that the night-time correction to the diameter is ~ 4.27 arc s and the daytime correction must be even larger, by ~ 2 –4 arc s, as discussed above. This means that the mean solar radius value of 964.5 arc s, determined by Picard during the deep Maunder minimum by micrometer measurements, must be corrected by at least 2.1 arc s, and probably by as much as 2 arc s more. This brings Picard's measurements down to between 962.4 and 960.4 arc s, which reduces

The cometary contribution to the oceans of primitive Earth

Christopher F. Chyba

Laboratory for Planetary Studies, Cornell University, Ithaca, New York 14853, USA

Recent compilations of the lunar impact record^{1,2}, combined with the mass-scaling law for crater diameters in the large-body (gravity-scaling) regime^{3,4}, allow an estimate of the total mass incident on the Moon during the period of heavy bombardment. By using only those craters and basins still extant, the calculation provides a lower limit which is independent of cometary flux models, extrapolations of the lunar impact record, or assumptions about the cometary mass spectrum. The results imply that the Earth would have acquired an exogenous ocean of water between ~4.5 and ~3.8 Gyr ago if comets comprised $\geq 10\%$ by mass of the impacting population.

In its classical formulation, the Oort cloud⁶ comprised 10^{11} – 10^{12} comets orbiting the Sun in a spherical shell extending from 10^4 to 10^5 AU. Perturbations by passing stars (and, we now know, by molecular clouds and galactic tides) would send some of these comets into the inner solar system where collisions with Earth could occur, leading to a terrestrial accretion of cometary material. An early estimate⁷ by Oró suggested Earth could have acquired 10^{14} – 10^{18} g in this way.

In comparison, the Earth's mass $M_{\oplus} = 5.98 \times 10^{27}$ g, and the terrestrial oceanic mass $M_{\text{oceans}} = 1.4 \times 10^{24}$ g H_2O . The total near-surface quota of water is greater, but quantifying it is difficult; typical estimates of water in the Earth's crust range from 2.0 to 3.3×10^{23} g (ref. 8). Of course much more water could be present deep in the interior. Here I use M_{oceans} as a reference, as it is the only certain figure. Taking comets to be ~50% water ice, Oró's result gives a cometary contribution of at most 10^{-7} to M_{oceans} . More recent estimates^{9–11} increased this number, but still allowed only $\leq 4\%$ of M_{oceans} to be cometary. (This conclusion would change dramatically if the interpretation of Frank *et al.*¹² of atmospheric dayglow emissions as a $\sim 10^{15}$ g per yr influx in ~10 metre comets were correct; such a flux would deliver $\sim M_{\text{oceans}}$ over the age of the Earth. But this interpretation has faced severe objections, including the facts that the proposed current flux is inconsistent by five orders of magnitude with Apollo lunar seismic data¹³, and that such comets should already have been detected optically¹⁴.)

In the past few years, difficulties with the classical Oort picture

and better models of planetary formation have suggested an inner Oort cloud with $\sim 10^{14}$ comets, beginning near the orbit of Neptune in the plane of the ecliptic, and slowly deforming outward into the classical Oort sphere¹⁵. Such models indicate that proto-Uranus and Neptune should have ejected from their vicinities planetesimals totalling many times their own final masses; the resulting flux of comets through the inner solar system early in its history may have been 10^4 – 10^5 times greater than that currently observed^{11,16,17}. These results have led to renewed speculation that comets may indeed have brought vast quantities of water to the primitive Earth^{16–19}. Here I use the lunar cratering record to place constraints on this cometary contribution. The results are summarized in Table 1.

A total of 33 known and 12 possible multi-ringed impact basins dating from the pre-Nectarian, Nectarian and Early Imbrian periods (extending from ~4.5 to ~3.8 Gyr ago) have now been discerned on the lunar surface¹. Cratering experiments suitable for simulating large impacts³ (the gravity scaling regime) relate crater diameter to impactor mass according to

$$M(D) = 0.089(\rho/v)^{6/5} g^{3/5} \delta^{-1/5} D^{18/5} \quad (1)$$

for an impactor of mass M , density δ , and incident velocity v , excavating a crater of diameter D on a body of density ρ and surface gravity g . This relation³ is derived from results for dry sand, scaled to competent rock according to data for half-buried explosive spheres. It is bracketed by equations such as those²⁰ based on the limiting cases of impact in dry sand and water.

For the Moon, $g = 161 \text{ cm s}^{-2}$ and the crustal density $\rho \sim 2.9 \text{ g cm}^{-3}$ (ref. 21). For the time being, we treat the impacting projectiles as Earth-crossing asteroids. The best value for v in equation (1) can be empirically determined from the orbital elements of the current population of these bodies; the root mean square lunar impact velocity (weighted according to probability of collision) for these objects has been found²² to be 16.1 km s^{-1} . Planetocentric bodies would have much lower velocities, implying more incident mass for craters of a given diameter.

Estimates of material density and void space for C- and S-type asteroids²³ suggest choosing $\delta \sim 2.0 \text{ g cm}^{-3}$. Investigations of terrestrial and lunar craters in the 3–100 km range indicate that wall collapse enlarges crater diameters by ~30% over their initial excavation limit²⁴. We incorporate this result by dividing the excavation diameter D in equation (1) by a crater collapse factor $c_f = 1.3$, although it is unclear that such enlargement is applicable in the case of ~1,000 km basins. The velocity appropriate to equation (1) is taken to scale as $\cos \theta$ (ref. 4), where θ is the impact incidence angle (although alternate angle-scaling relationships have been proposed)^{24,25}. The average

Table 1 Cometary contribution to the oceans of primitive Earth

Method of calculation*	Sum impacting mass implied by all extant basins and craters dating from 3.8 to 4.5 Gyr				Sum Early Imbrian and Nectarian impacting mass, and integrate back to 4.5 Gyr		
	Innermost		Topographic		Innermost	Topographic	Topographic
Basin ring taken to represent the limit of excavation	Known†	Known + possible	Known	Known + possible	Known‡	Known‡	Known§
Basins included in calculation	Known†	Known + possible	Known	Known + possible	Known‡	Known‡	Known§
Impacting mass implied by basins (g)	2.3×10^{22}	5.9×10^{22}	4.2×10^{23}	1.1×10^{24}	8.6×10^{22}	1.6×10^{24}	8.6×10^{24}
Impacting mass implied by craters (g)	5.7×10^{21}	5.7×10^{21}	5.7×10^{21}	5.7×10^{21}	1.7×10^{21}	3.2×10^{22}	1.8×10^{23}
% total impacting mass due to craters	19.9	8.8	1.4	0.5	2.0	2.0	2.0
Total mass delivered to Earth (g)	7.6×10^{23}	1.7×10^{24}	1.1×10^{25}	2.9×10^{25}	2.3×10^{24}	4.3×10^{25}	2.3×10^{26}
Total H_2O delivered if impactors 10% cometary (in units of M_{oceans})	0.03	0.06	0.39	1.0	0.08	1.5	8.2

* All calculations assume asteroidal parameters.

† This column gives a lower limit to the total mass delivered.

‡ Nectaris event dated at 4.1 Gyr.

§ Nectaris event dated at 3.92 Gyr.

|| Calculation neglects H_2O contributed by asteroids.

incidence angle over a hemisphere is $\langle \theta \rangle = 57.3^\circ$. Equation (1) then becomes

$$M(D) = 2.70 \left(\frac{2.0 \text{ g cm}^{-3}}{\delta} \right)^{1/5} \left(\frac{16.1 \text{ km s}^{-1}}{v} \right)^{6/5} \times \left(\frac{0.540}{\cos \theta} \right)^{6/5} \left(\frac{1.3}{c_f} \right)^{18/5} D^{18/5} \quad (2)$$

with D in metres and M in grams.

We now apply equation (2) to the observed lunar basins to determine the masses of the excavating impactors. Immediately we confront the controversy over which ring of such multi-ringed structures should be taken as the boundary of excavation to give D in equation (2), some investigators favouring the innermost, and some the main, topographic, ring. In the former view the outer rings form by inward collapse towards the recently excavated crater; in the latter the inner rings are due either to floor uplifts or the "freezing in" of ripples in the fluid crust¹. I find the evidence for the topographic ring compelling because the differences between smaller basins (such as Schrödinger) and large, single-ringed craters would otherwise be difficult to explain¹, and also because of the well-known morphological and cratering differences observed across the boundary of the main ring²⁶. Nevertheless, all calculations are performed for both choices of excavation limit. Results should therefore bracket the correct value. For each excavation limit, two cases are considered. First we include only those 33 basins whose existence is certain (Table 1, columns labelled 'known'), then add 12 basins whose existence is controversial ('known + possible'). Thus we find, for example, that choosing the topographic rim means that the impactors responsible for the 33 known basins had a combined mass $M_{\text{basins}} \sim 4.2 \times 10^{23}$ g.

How much mass was incident in the form of projectiles too small to form multi-ringed basins? Crater counts from the Nectaris basin imply that 1,700 craters 20 km and larger formed on the entire Moon during the Nectarian period¹, whereas counts from other basins give a Nectarian large-crater frequency three times higher than the Imbrian, and one third that of the pre-Nectarian². Thus a lower limit to the number of such craters formed during these periods is $N \sim 7,400$. This is a lower limit as highland data show even higher crater frequencies², and could be a gross underestimate should these data represent saturation rather than production distributions. The observed number of lunar craters larger than diameter D falls off roughly as D^{-2} (ref. 5), so that the differential number of craters at a particular diameter can be written $dN(D) = \alpha D^{-3} dD$. The proportionality constant α can be fixed by integration for a known N . Equation (2), valid for lunar craters with D above ~ 1 km (ref. 3), can then be combined with this to give the total mass M_{craters} contributed by the N small impactors

$$M_{\text{craters}} = \int_{D_1}^{D_2} M(D) dN(D) \quad (3)$$

With $D_1 = 20$ km and $D_2 = 300$ km (the lower limit of basin diameters), $M_{\text{craters}} = 5.7 \times 10^{21}$ g, representing $\sim 1.4\%$ of the total mass $M_{\text{impacts}} = M_{\text{basins}} + M_{\text{craters}}$ incident on the Moon between ~ 4.5 and ~ 3.8 Gyr ago. $M_{\text{impacts}} \sim 0.6\%$ the current lunar mass. More recent cratering contributes negligibly to this sum⁵. Depending on how many basins and craters were overlain and obliterated, it could be a considerable underestimate.

For the Earth-crossing asteroids, the mean probability per year per asteroid for a collision with the Earth is 26.3 times greater than that for a collision with the Moon²². Thus $\sim 26 M_{\text{impacts}} = 1.1 \times 10^{25}$ g should have impacted Earth between 4.5 and 3.8 Gyr ago. Now consider the consequences if some significant fraction of these impactors were comets. Water vaporized in a cometary impact will simply condense and rain out. Taking comets to be 50% H_2O , and assuming for definiteness $\sim 10\%$ by mass of the impactors to have been cometary (we will discuss such estimates below), Earth should have acquired $\sim 0.4 M_{\text{oceans}}$ of H_2O . As the mass delivered cannot be less than

that calculated by using the innermost rings of only the 33 known basins, column 1 of Table 1 provides a firm lower limit to such calculations.

Do impacting comets arrive significantly devolatilized? The statistics of cometary disappearances suggest that short period (SP) comets have mean lifetimes $\sim (100-300)q^{1/2}$ perihelion passages (where q is the perihelion distance)²⁷, although this may be a considerable underestimate: dynamic calculations imply that comet Halley has already made $\sim 2,000$ perihelion passages as an SP comet²⁸. The comets considered here, those creating craters ≥ 20 km, have radii ≥ 1.1 km. If volatile loss is controlled by water ice sublimation, such comets with $q = 1$ AU would require thousands of perihelion passages to sublime completely, and these sublimation lifetimes are probably much further prolonged by the formation of an involatile nuclear crust²⁸. In any case, slow, cumulative devolatilization appears not to be the lifetime-limiting process; large comets end their lives before they are significantly devolatilized. Observations of comet Halley²⁹ indicate a gas/dust mass production ratio ~ 3.3 , with $\sim 80\%$ of the gas molecules being H_2O . Thus Halley may well remain more than 50% H_2O by mass. In addition, some unknown fraction of colliding 'asteroids' may have originated as extinct cometary nuclei²⁴, a source of volatiles not included here.

Impact modelling³⁰ shows that planets accrete more material than they lose in ejecta for impacts of 1.0 g cm^{-3} comets at 15 km s^{-1} , provided $v_{\text{esc}} = 2.5 \text{ km s}^{-1}$ or larger. Thus the Moon ($v_{\text{esc}} = 2.4 \text{ km s}^{-1}$) loses about as much as it gains. At v_{esc} for Earth (11.2 km s^{-1}), however, virtually all material is retained. Of much greater importance (and controversy) is the question of erosion of the Earth's atmosphere due to large impacts. Walker³¹ has found that atmospheric impact erosion does not occur for impacts of any velocity on bodies with $v_{\text{esc}} \geq 10 \text{ km s}^{-1}$. Thus, whereas atmospheric erosion is possible for Mars or the Moon, Earth accretes virtually all impacting mass. By contrast, however, it has been suggested³² that hydrodynamic acceleration of completely vaporized material, following the impact of a body of mass $\geq 10^{30}$ erg (that is, the impacts of relevance to this work), may indeed cause terrestrial atmospheric loss. If this latter conclusion is correct, a significant fraction of cometary water delivered to Earth might thus be lost.

Repeating the preceding calculation for the case where the innermost ring is taken to represent the limit of excavation is difficult, as many of the oldest (pre-Nectarian) basins are significantly degraded, so that inner (and perhaps some outer) rings cannot be discerned. An examination of the 14 'known' basins, dating from the Early Imbrian and Nectarian periods (~ 3.8 to ~ 4.1 Gyr) (refs 1, 11), shows that the diameters of these basins' innermost rings are, on the average, 0.44 times as large as those of the topographic rings. To err on the side of conservative mass estimates, I therefore always treat the largest observed basin ring as the 'topographic' ring, and compute the impacting mass implied by the (presumed unobserved) innermost ring by multiplying the topographic result given by equation (2) by $(0.44)^{18/5} = 0.054$.

The strength of the method of calculation thus far employed is its independence of cometary flux models and extrapolations of the lunar cratering chronology. Its weakness is that the large number of 'possible' pre-Nectarian basins demonstrates that the major portion of the ancient cratering record may be altogether lost. For example, including the 12 'possible' basins (see Table 1) in the calculation for the topographic ring increases the mass delivered to the Earth by a factor of ~ 2.5 , so that $\sim 1 M_{\text{oceans}}$ is delivered to Earth if 10% of the impactors are cometary. Perhaps even this is an underestimate.

This problem can be avoided by using only the 14 'known' basins dating from the Early Imbrian and Nectarian periods. Coming at the end of heavy bombardment, these basins are better preserved and should more completely represent their periods' impact record. We can then integrate back to 4.5 Gyr

using the observed 100 Myr half-life of the lunar cratering chronology⁵, scaling by the mass implied by the Early Imbrian and Nectarian basin and crater record. This calculation replaces one uncertainty (obliteration of impact features) with three others, namely the extrapolation of the lunar cratering chronology, the applicability of this chronology to basin-forming impactors, and the dating of the basins themselves. For example, ³⁹Ar/⁴⁰Ar analysis of Apollo 16 lunar samples yields two possible ages for the Nectaris event, 4.1 and 3.92 Gyr (refs 1, 11); the different ages change the estimate of the mass delivered to the Earth by a factor ~5 (see Table 1). A more serious objection is that the sudden drop-off of the lunar basin-forming record at ~3.8 Gyr may mean that the late heavy bombardment is a pulse superimposed on a smaller, more slowly decaying flux³³. If this is the case, then the preceding calculation overestimates the actual flux. Nevertheless, as seen in Table 1, this latter, more speculative calculation yields results in order-of-magnitude agreement with those obtained by the entirely distinct, more conservative method of simply summing the mass implied by those basins and craters which are still extant.

What independent checks are available for these results? At least four should be considered:

(1) **The H₂O inventory of the other terrestrial planets.** Comets should have delivered water to the other terrestrial bodies as well. Study of lunar samples is consistent with no measurable H₂O content. But on the Moon, most impacting cometary H₂O would be lost in ejecta, with the remainder susceptible to atmospheric blow-off, photodissociation, and destruction by the solar wind. Several sources of lunar H₂O during the past 2 Gyr (including recent cometary impacts) probably each supplied 10¹⁶–10¹⁷ g of water; its complete absence testifies to the high efficiency of these destructive mechanisms (although some water may remain in polar cold traps)³⁴. Mars, with a smaller cross-section and a smaller v_{esc} (5.0 km s⁻¹) than Earth, would be expected to acquire and retain less impact-delivered H₂O. Many martian landforms suggest the former presence of considerable ground ice or water, and there appears at present to be ice-rich terrain poleward of ±30° latitude³⁵. Finally, the deuterium-hydrogen (D/H) ratio in the Venus atmosphere suggests the former presence of much water on Venus³⁶. These observations are consistent with (but do not independently support) a significant cometary source of H₂O.

(2) **Morphological and compositional characteristics of lunar impacts.** Can we distinguish between asteroidal and cometary lunar impact features? Certain magnetized lunar swirl patterns have been interpreted³⁷ as results of recent (≤100 Myr) cometary impacts. But it has been argued³⁸ that the features are in fact the antipodal effects of ancient basin-forming impact events, so that no cometary origin needs to be invoked. In any case, the associated craters do not appear to be anomalous. Comets have been typically thought of as 0.1–10 km bodies, although recent imaging of comet Halley has shown it to be darker and larger than commonly expected (although some authors³⁹ had anticipated C-type asteroidal albedos for SP comet nuclei), with major and minor axes of 15 and 7–10 km (ref. 40). A Halley-sized comet, assuming a typical SP comet lunar impact velocity ~26.7 km s⁻¹ (ref. 41) and a density $\delta \sim 1.0 \text{ g cm}^{-3}$, would leave behind a lunar crater ~80 km across. To excavate a Serenitatis-sized basin (740 km, representing ~1% of the mass delivered to the Moon by the 33 known basins and craters) would require a comet of radius ~80 km. But the Great Comet of 1729 (a naked-eye object with perihelion ~4 AU) (ref. 42) and the trans-Saturnian object 2060 Chiron (radius ~200 km for a Halley-like albedo of 0.03) (ref. 43) may provide examples of such comets. Enrichment of siderophile trace elements in some lunar highland breccias indicates sampling of basin-forming projectile remnants, although interpretation remains controversial¹¹. Ternary plots show that, whereas the siderophile composition of most of the nine putatively-identified impactors is distinct from the composition of carbonaceous chondrites (the meteoritic class

most likely to resemble comets), at least one (represented by the group 5L clasts) does have C1 proportions⁴⁴. The C1 are the most primitive and volatile-rich (presumably the most 'comet-like') of the chondrites. All impactor remnants, however, are depleted in volatiles by factors of 2–100 relative to carbonaceous chondrites⁴⁵. But this is a poor diagnostic, as these elements would be preferentially lost at the high post-impact temperatures.

(3) **Isotopic data.** It has been argued⁴⁶ that hydrogen and carbon isotopic data require an exogenous source of surface volatiles for Earth and Mars. It is suggestive that the oceanic D/H ratio (1.6×10^{-4}) and that of comet Halley ($0.6\text{--}4.8 \times 10^{-4}$) are both an order of magnitude above cosmic abundance ($2.0 \pm 1.0 \times 10^{-5}$) (ref. 47). Some authors have also interpreted findings that juvenile water from Earth's interior contains 15% less deuterium than the oceans⁴⁸ as evidence for an extraterrestrial source of H₂O (refs 46, 48).

(4) **The cratering record throughout the Solar System.** Estimates of current asteroidal and cometary cratering rates suggest that comets account for 10–30% of the recent production of terrestrial impact craters ≥10 km (ref. 24), although the inclusion of 'extinct' comets⁴⁹ or possible comet showers⁵⁰ could raise this fraction to above 50%. How to extrapolate this result back to the early Solar System is not clear, however. A recent model of outer planet formation implies a cometary flux through the inner Solar System declining with an approximate 150 Myr half-life between 4.5 and 3.8 Gyr (before levelling off to its current relatively low value), possibly delivering as much as ~100 M_{oceans} to the Earth¹⁶. Any such enormous influx of water would have to be largely lost; possible mechanisms would include atmospheric erosion, photochemical dissociation, and incorporation into the Earth's core (D. Stevenson, personal communication). Other models for cometary flux and mass spectrum may be inconsistent with the delivery of so much H₂O (ref. 11). The empirical, model-independent approach taken here serves as an observational diagnostic for such theoretical models.

Crater diameter-density distributions for the terrestrial planets imply a common heliocentric impactor population for the heavy bombardment in the inner Solar System²⁵. But the great disparities between the cratering records of the Moon and those of some outer planet satellites strongly suggests that comets were not the major component of this population²⁵. Any scenario in which Earth receives many tens of oceans of water or more must explain this contradiction between the cratering record implied by such an influx and that which is in fact observed. Admittedly, as long as only the lunar cratering record can be assigned absolute ages, such an objection may remain indecisive. For the results found here, this diagnostic is in any case not very useful as Earth may have acquired ~1 M_{oceans} if only ~10% of the impactors were comets. That is, Earth could owe much of its oceans to comets, but the cometary signature in the cratering record could still be utterly dominated by asteroidal or planetocentric impactors. This is not necessarily to deny the usual account of terrestrial volatile origin through outgassing; undoubtedly the two processes were complementary. But despite the uncertainties in the preceding calculations, it appears that an extraterrestrial source for a significant fraction of the Earth's oceans can no longer be rejected.

I am indebted to Joseph Burns, Carl Sagan, Paul Weissman, Eugene Shoemaker, David Stevenson and George Wetherill for their critical reviews of this manuscript, and to Paul Helfenstein and Gerhard Neukum for helpful discussions. This work was supported in part by NASA.

Received 24 July; accepted 22 September 1987.

1. Wilhelms, D. E. in *The Geology of the Terrestrial Planets* (ed. Carr, M. H.) 107–205 (NASA SP-469, 1984).
2. Neukum, G. & Wilhelms, D. E. *Lunar planet. Sci.* **13**, 590–591 (1982).
3. Schmidt, R. M. & Holsapple, K. A. in *Geological Implications of Impacts of Large Asteroids and Comets on the Earth* (eds Silver, L. T. & Schultz, P. H.) 93–101 (Geol. Soc. Am. SP-190, 1982).

4. Chapman, C. R. & McKinnon, W. B. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 492-581 (Univ. Arizona, 1986).
5. Neukum, G. thesis, Univ. München (1983).
6. Oort, J. H. *Bull. astr. Insts. Neth.* **11**, 91-110 (1950).
7. Oró, J. *Nature* **190**, 389-390 (1961).
8. Carr, M. H. in *The Geology of the Terrestrial Planets* (ed. Carr, M. H.), 79-105 (NASA SP-469, 1984).
9. Pollack, J. B. & Yung, Y. L. *A. Rev. Earth planet. Sci.* **8**, 425-487 (1980).
10. Oró, J., Holzer, G. & Lazcano-Araujo, A. in *COSPAR: Life Sci. Space Res.* **18**, 67-82 (1980).
11. Wetherill, G. W. *Multi-ring Basins, Proc. Lunar Planet. Sci. Conf.* **12A**, 1-18 (1981).
12. Frank, L. A., Sigwarth, J. B. & Craven, J. D. *Geophys. Res. Lett.* **13**, 307-310 (1986).
13. Davis, P. M. *Geophys. Res. Lett.* **13**, 1181-1183 (1986).
14. Soter, S. *Geophys. Res. Lett.* **14**, 162-163 (1987).
15. Weissman, P. R. *Space Sci. Rev.* **41**, 299-349 (1985).
16. Fernandez, J. A. & Ip, W.-H. *Icarus* **54**, 377-387 (1983).
17. Shoemaker, E. M. & Wolfe, R. F. *Lunar planet. Sci.* **15**, 780-781 (1984).
18. Whipple, F. L. *Mém. Soc. Sci. Liège* **9**, 101-111 (1976).
19. Chyba, C. F. *Bull. Am. Astr. Soc.* **18**, 782 (1986).
20. Vickery, A. M. *Icarus* **67**, 224-236 (1986).
21. Haines, E. L. & Metzger, A. E. *Proc. Lunar Planet. Sci. Conf.* **11**, 689-718 (1980).
22. Shoemaker, E. M. & Wolfe, R. F. *Lunar Planet. Sci.* **18**, 918-919 (1987).
23. Shoemaker, E. M. *et al.* in *Asteroids* (ed. Gehrels, T.) 253-282 (University of Arizona, Tucson, 1979).
24. Shoemaker, E. M. *A. Rev. Earth planet. Sci.* **11**, 461-494 (1983).
25. Strom, R. G. *Icarus* **70**, 517-535 (1987).
26. Baldwin, R. B. *Multi-ring Basins, Proc. Lunar Planet. Sci. Conf.* **12A**, 275-288 (1981).
27. Krésák, L. in *Dynamics of Comets: Their Origin and Evolution* (eds Carusi, A. & Valsecchi, G. B.) 279-302 (Reidel, Boston, 1985).
28. Weissman, P. R. in *Diversity and Similarity of Comets* (ESA SP-278, in the press).
29. Whipple, F. L. in *20th ESLAB symposium on the exploration of Halley's comet 2*, 281-288 (ESA SP-250, 1986).
30. O'Keefe, J. D. & Ahrens, T. J. *J. Geophys. Res.* **87**, 6668-6680 (1982).
31. Walker, J. C. G. *Icarus* **68**, 87-98 (1986).
32. Ahrens, T. J. & O'Keefe, J. D. *Lunar Planet. Sci.* **18**, 9-10 (1987).
33. Wetherill, G. W. *Proc. Lunar Sci. Conf.* **8**, 1-16 (1977).
34. Arnold, J. R. *J. Geophys. Res.* **84**, 5659-5668 (1979).
35. Carr, M. H. *Nature* **326**, 30-35 (1987).
36. Donahue, T. M. *et al.* *Science* **216**, 630-633 (1982).
37. Schultz, P. H. & Srnka, L. J. *Nature* **284**, 22-26 (1980).
38. Hood, L. L. *Lunar Planet. Sci.* **18**, 433-434 (1987).
39. Shoemaker, E. M. & Wolfe, R. F. in *Satellites of Jupiter* (ed. Morrison, D.) 277-339 (Univ. Arizona, 1982).
40. Keller, H. U. *et al.* *Nature* **321**, 320-326 (1986).
41. Weissman, P. R. in *Geological Implications of Impacts of Large Asteroids and Comets on the Earth* (eds Silver, L. T. & Schultz, P. H.) 15-24 (Geol. Soc. Am. SP-190, 1982).
42. Kronk, G. W. *Comets: A Descriptive Catalogue* (Enslow, Hillside, New Jersey, 1984).
43. Kowal, C. T. in *Asteroids* (ed. Gehrels, T.) 436-439 (Univ. Arizona, 1979).
44. Hertogen, J. *et al.* *Proc. Lunar Sci. Conf.* **8**, 17-45 (1977).
45. Gros, J. *et al.* *Proc. Lunar Sci. Conf.* **7**, 2403-2425 (1976).
46. Anders, E. & Owen, T. *Science* **198**, 453-465 (1977).
47. Eberhardt, P. *et al.* in *24th ESLAB Symp. Exploration of Halley's Comet 1*, 539-541 (ESA SP-250, 1986).
48. Kokubu, N., Mayeda, T. & Urey, H. *Geochim. cosmochim. Acta* **21**, 247-256 (1961).
49. Shoemaker, E. M. & Wolfe, R. F. *Lunar planet. Sci.* **17**, 799-780 (1986).
50. Shoemaker, E. M. & Wolfe, R. F. in *The Galaxy and the Solar System* (eds Smolouchowski, R. *et al.*) 338-386 (1986).

Electric field-induced flame convection in the absence of gravity

F. B. Carleton & F. J. Weinberg

Department of Chemical Engineering and Chemical Technology,
Imperial College, London SW7 2BY, UK

Although chemical fuels uniquely provide a safe way of carrying large amounts of energy within a small mass and volume, their combustion in the absence of gravity raises major problems. Whenever flames are used for heating specific objects under gravity and in an oxygen-containing atmosphere, we rely on natural or forced convection to replenish reactants and direct the hot products. Forced convection requires a compressor or keeping fuel under pressure in cylinders, and thus entails a considerable weight penalty. Natural convection is not available under micro-gravity conditions, making heat transfer from flames difficult to achieve. Indeed, diffusion flames will tend to become spherical and will eventually extinguish as a result of blanketing by their own products. We report here an alternative method of controlling flames in zero gravity based on the application of electric fields, which can be achieved using compact, lightweight equipment.

It is well known (see, for example, ref. 1) that relatively large concentrations of free charge occur in the reaction zones of flames due to chemi-ionization. In hydrocarbon flames, these concentrations can exceed the steady-state (Saha) equilibrium values by factors of the order of 10^4 . The negative charge carriers are largely electrons, whereas positive charge is carried by molecular ions of which H_3O^+ predominates initially. In electric fields well below the breakdown value ($<10 \text{ kV cm}^{-1}$), such ions do not continue to accelerate beyond their terminal velocity, v , as they have a constant mobility, k , given by

$$v = kE$$

where E is the local electric field strength. Consequently, they exercise a drag upon the neutral gas resulting in a body force, F , per unit volume given by

$$F = neE$$

where ne is the space charge contained in that volume. This applies for fields large enough to ensure that the charge clouds in the regions between flame and each electrode are unipolar. The general expression is $F = Ee(n_+ - n_-)$ but, in the presence of an appreciable E , charge carriers of opposite polarities co-exist only in the thin chemi-ionization zone. Under these condi-

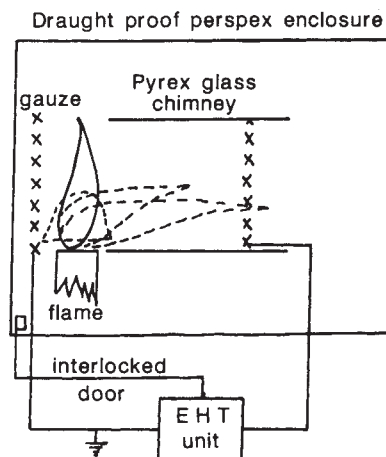


Fig. 1 Electro-gas dynamic chimney.

tions the local current density, j , is given by

$$j = nekE$$

so that

$$F = j/k$$

The distribution of body forces for each particular geometry leads to an 'ionic' or 'Chattock'² wind which can readily be used to break down and secondary ionization are encountered.

For appreciable path lengths in cold air, the initial negative charge carriers which are electrons, attach to produce negative ions. Thereafter positive and negative charge carriers have similar mobilities so that gas flows from the flame to each electrode in a symmetrical system. Unidirectional flows are induced most readily by placing the positive electrode in close proximity to the flame. The body force due to natural convection reaches a maximum value in earth gravity, g , of ρg , where ρ is the ambient air density, and it has been shown³ that the body force which can be made to act on flames by the application of electric fields can reach 800 times that value before limitations due to breakdown and secondary ionization are encountered.

The gravitational and electrical body forces are otherwise analogous in every way and an 'electro-gas dynamic chimney', such as that shown in Fig. 1, can be constructed readily so as to provide a 'draught', just as a convective chimney would. For

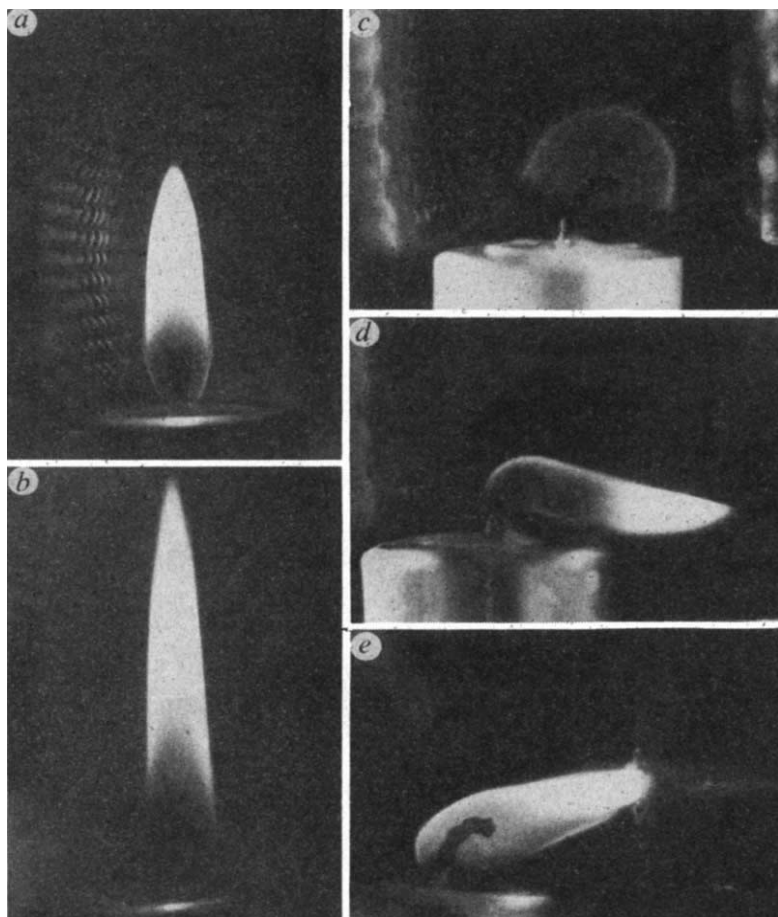


Fig. 2 Flame shapes under varying conditions of gravity and field strength. *a*, Conditions of 1g and zero field; *b*, 2g, zero field; *c*, zero g, zero field; *d-e*, zero g, increasing field strength.

a one-dimensional field, if N is the number of g to be simulated, a current J is drawn, where

$$J = ANk\rho g$$

and A is the cross-sectional area. The maximum gas velocity which can be induced in such a one-stage ion pump has been shown^{1,3} to be $\sim 5.5 \text{ ms}^{-1}$ which is generally much in excess of combustion requirements. Although the potentials are large, the current drawn and the energy dissipated electrically are insignificant. The maximum current that can be drawn before electrical breakdown sets in is only 0.25 mA cm^{-2} of flame front and the corresponding power density is $< 0.1 \text{ W cm}^{-2}$.

The electronic high-tension units used here have become progressively smaller, lighter and cheaper. Recently a device has been developed⁴ for the "Electrodyne" hand-held electrostatic insecticide sprayer for Third World use that features a high voltage supply powered by torch batteries, small enough to fit into the handle of the unit. With simple equipment of this kind, any desired burner flow can be induced and hot flame gases can be directed specifically to the surfaces to be heated.

Practical tests of the above concept were sponsored by the European Space Agency and tested in KC135 parabolic aircraft flights from NASA, Houston. These allow for repeated micro-gravity experiments of duration not exceeding 30 s, which is not long enough to obtain a steady state for calorimetric measurements, but is more than adequate for photographic recording of flow patterns. For reasons of safety, the work had to be carried out with candles (in which liquid fuel of $\sim 43 \text{ kJ g}^{-1}$ is released gradually by the melting of the solid wax). The apparatus used is illustrated in Fig. 1 and the flame behaviour was recorded by a Minolta programmable 7000 camera, which automatically compensates for changing light conditions and minimizes

experimental difficulties associated with free-fall conditions.

Figure 2a shows the flame under normal 1g conditions. The effect of doubling g , during the period when the aircraft pulls out of the parabola, is shown in Fig. 2b; Fig. 2c shows the flame in the absence of gravity without an applied electric field. It is virtually spherical (except for the region quenched by the proximity of the fuel). Although flames did not extinguish completely during the period of micro-gravity available, they became so non-luminous as to be invisible to the naked eye. (Some impression of the extent of the camera's compensation for that lack of light may be gathered from the visibility of the background, a dark cloth.) Figure 2d, e shows the effect of increasing electric fields applied to the system shown in Fig. 1.

Although the heat source had to be confined to a small diffusion flame, the power flux, calculated from the rate of the steady-state fuel consumption and the cross-sectional area of the hot streamer in Fig. 2e, at $\sim 300 \text{ W cm}^{-2}$, is quite respectable because of the focusing of the streamlines by the electric field onto the region to be heated. There is little doubt that this principle, using simple and lightweight equipment, could be applied to provide intense heating of small areas, making economic use of the oxygen available in the working environment, and using liquid or solid fuels requiring no pressurized containment.

We are indebted to the European Space Agency for supporting part of this work, and to NASA for the flight opportunity.

Received 9 July; accepted 6 October 1987.

1. Lawton, J. & Weinberg, F. J. in *Electrical Aspects of Combustion* (Clarendon, Oxford, 1969).
2. Chattock, K. P. *Phil. Mag.* **48**, 401-420 (1899).
3. Lawton, J., Mayo, P. J. & Weinberg, F. J. *Proc. R. Soc. A* **303**, 275-298 (1968).
4. Coffee, R. A. *Outlook on Agriculture* **10**, 350-356 (1981).

Specific heat anomaly at 220 K connected with superconductivity at 90 K in ceramic $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$

T. Lægrend*, K. Fossheim*, E. Sandvold*† & S. Julsrud‡

* Department of Physics and Mathematics, The Norwegian Institute of Technology, N-7034 Trondheim-NTH, Norway

‡ Institute of Inorganic Chemistry, The Norwegian Institute of Technology, N-7034 Trondheim-NTH, Norway

Extraordinary efforts have recently been put into trying to understand the new class of high-temperature superconducting copper oxides discovered by Bednorz and Müller¹. Here, we present high-resolution specific heat measurements on four different samples of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (YBCO) which reveal a new anomaly at 220 K (see also ref. 2). The strength of the anomaly as measured by its height is related to the size of the jump at the superconducting transition at 90 K. Specific heat measurements are sensitive and reliable volume probes of ordering phenomena in the bulk of solids; our results therefore establish that there are ordering phenomena taking place in YBCO at ~ 220 K. These findings may be of importance for understanding reports of oxygen ordering^{3,5} and of various unstable superconducting transitions in the range from 150 K (ref. 6) to well above 200 K (ref. 7).

Samples UK1, UK3, UK3NEG and B39/2 were all prepared from Y_2O_3 , BaCO_3 and CuO (pro analysis), milled together, pressed into pellets and fired at 950°C in air². UK1 was then furnace-cooled in air, taking ~ 7 h to reach room temperature. UK3 was milled and pressed a second time, fired at 950°C , cooled to room temperature, and finally held at 450°C for 10 h in flowing oxygen. UK3NEG was cut from the UK3 batch and treated at 450°C for 10 h in flowing nitrogen gas. B39/2 was fired a second time at 950°C for 16 h, cooled at $1.5^\circ\text{C min}^{-1}$ to 650°C and held for 8 h, and finally cooled at $1.5^\circ\text{C min}^{-1}$ to room temperature, all in flowing oxygen. Powder X-ray analysis showed that all samples were single-phase YBCO, but UK1 contained small traces of starting material. The samples exhibited levitation and zero resistivity below the 90-K superconducting transition, except for UK3NEG, from which enough oxygen was removed to prevent the superconducting transition above liquid-nitrogen temperature⁸.

The relative change in specific heat was measured using an a.c. technique⁹, and the data were calibrated using the value of C_p at 100 K from ref. 10. The heating frequency used was 1 Hz, and the amplitude of the temperature oscillations generated was ~ 50 mK. The scatter in our specific heat data is less than 1 part in 1,000 and the temperature at each measuring point was stabilized to better than 5 mK. This was achieved by regulating the sample-holder temperature for 5 min before measurement, thus giving ample time for any ordering process to occur. The temperature difference between each measuring point was ~ 0.23 K. The phase and amplitude of the a.c. signal was measured using different frequencies between 0.5 and 2.5 Hz, thus providing a check⁹ that the thickness of the sample and the measuring frequency had been chosen such that thermal quasi-equilibrium was obtained across the sample to the thermocouples. At the same time thermal relaxation to the surroundings was negligible. Thus, the choice of experimental parameters ensures that the thin-sample limit applies: neither the finite value nor the observed variations of the thermal diffusivity¹¹⁻¹³ will present any problem. In fact, the 20% variation in thermal diffusivity typically found over the temperature range of our measurements can be shown to change our measured value of the relative specific heat by only 4%.

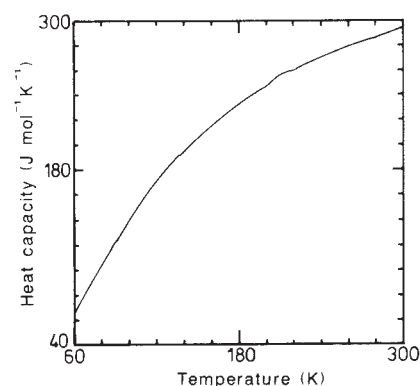


Fig. 1 Specific heat of sample UK3 between 60 and 300 K. 991 Data points are shown. The anomaly at the superconducting transition at 90 K is too small to be visible in this plot, but the anomaly at 220 K can be clearly seen.

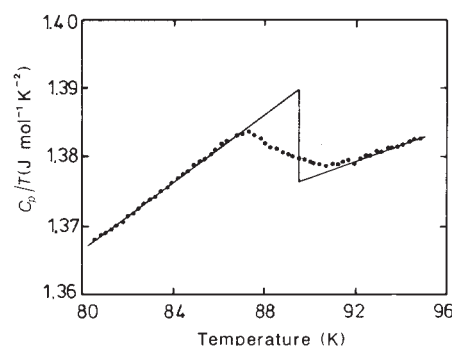


Fig. 2 The specific heat anomaly of sample UK1 near the superconducting transition, plotted as C_p/T versus T . The magnitude of the jump is determined as indicated by the solid line.

Figure 1 shows the specific heat of sample UK3 over the whole temperature range measured. The anomaly at ~ 220 K is clearly visible, but the anomaly at the superconducting transition at 90 K is not easily seen on this scale, owing to the large lattice background. The resolution of the experiment is good enough, however, for the 90-K anomaly to be well resolved when analysed further.

To study the specific heat jump at the 90-K superconducting transition, we found it useful to plot the data as C_p/T versus T ; Figure 2 shows an example of such a plot for sample UK1. Table 1 lists the results of the analysis of the four different samples, derived from several measurements on each sample. T_c is determined from the midpoint of the resistive transition. We believe that the large spread in ΔC_p for the different samples is caused by varying volume fractions of optimally superconducting material, related to differences in oxygen content and distribution imposed by the different preparation procedures. A thorough analysis and discussion of the results at the 90-K superconducting transition can be found in ref. 2. The thermodynamic consistency of specific heat and ultrasonic measurements in YBCO and LBCO ($\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$) has also been reviewed¹⁴.

A systematic analysis of the anomaly at 220 K was achieved by subtracting from the data a quadratic background fitted to points taken in the temperature ranges 180–190 K and 230–245 K. The resulting anomaly for sample B39/2 is shown in Fig. 3. This procedure was carried out for all samples and the findings are summarized in Table 1. $\Delta C_{p,h}$ is determined as the maximum of the specific heat anomaly. This need not be a well-defined quantity if the anomaly is caused by a transition with fluctuation contributions, but seems to be the most reasonable choice for a quantitative comparison between different samples. Note the connection in the magnitude of the specific heat anomalies at 90 and 220 K; although the ratios $\Delta C_{p,h}/\Delta C_p$ are not perfectly

† Present address: ELAB, N-7034 Trondheim, Norway.

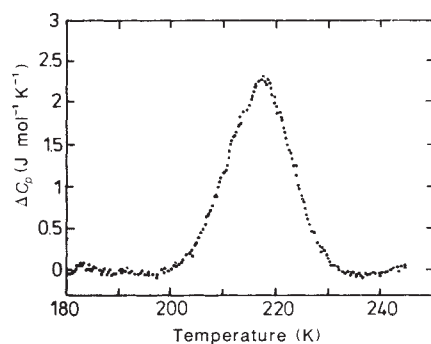


Fig. 3 The high-temperature specific heat anomaly of sample B39/2 near 220 K, obtained by subtracting a quadratic background.

equal for the different samples, the jump magnitudes span a factor of 13, while the ratio changes by only a factor of two. Interestingly, the strong suppression of the superconducting anomaly in UK3NEG is accompanied by a simultaneous suppression of the 220 K anomaly in that sample. This is a strong indication that the same or related mechanisms are at play leading to the observed anomalies at 90 and 220 K.

We discuss now some possible explanations for our observations. The effect of oxygen content on the transition temperature is well known, and the connection between the 90-K and 220-K anomalies suggests that ordering in the oxygen system may be responsible for the 220-K anomaly. Recent neutron diffraction experiments¹⁵ and high-resolution transmission electron microscopy⁴ do in fact seem to show long-range order in the oxygen vacancies. The reported achievement of superconducting transition temperatures up to 159 K by "phase healing" thermal cycling⁶ may also reflect ordering of the oxygen system at temperatures near 220 K. Furthermore, it has been proposed that the microscopic oxygen configuration affects the magnetic susceptibility through its influence on the copper ions, and that there exists a single-phase 60-K bulk superconductor characterized by an ordered array of oxygen vacancies³. These vacancies are known to exist almost exclusively in the copper-oxygen chains in YBCO, and could produce separated regions of Cu^{2+} and Cu^{3+} ions^{16,5}. Oxygen ordering is made responsible for the superconductivity in $\text{La}(\text{Ba}_{2-x}\text{La}_x)_2\text{Cu}_3\text{O}_{7-\delta}$ ¹⁷ and for the very high but unstable superconducting transition temperature of 210 K in a mixed-phase yttrium barium copper oxide⁷. The tetragonal polymorph of YBCO, which is characterized by disorder of the oxygen in the a - b plane, can be obtained in samples quenched from above the phase transition occurring at $\sim 700^\circ\text{C}$. Such samples have strongly depressed transition temperatures or do not show any transition at all. From these observations it is inferred that either the ordering of oxygen or the presence of Cu^{3+} ions is important for the superconducting transition in YBCO¹⁸. The degree of order is also important in a recently proposed fracton-mediated electron-electron coupling mechanism¹⁹ and in defect-enhanced electron-phonon coupling²⁰.

Some ordering involving the different valence states of copper could also be an explanation for our observed anomaly, and a small decrease of the susceptibility near 240 K has been reported²¹. Ultrasonic and flexural-mode measurements of the elastic properties of YBCO have indicated a possible structural phase transition^{22,23} at ~ 230 K, which should then influence the specific heat. A theory considering arrays of ultra-small Josephson junctions can predict two separate superconducting transitions²⁴ and could lead to two specific heat anomalies. Nuclear quadrupole resonance measurements have been analysed as reflecting two gaps, from which transition temperatures of 60 K and 200 K are inferred, with the higher temperature associated with the chains²⁵. Finally, a possible contribution to the specific heat from contaminations of the samples by other phases has to be considered. In view of the high quality of the samples as determined by X-ray analysis and the quite large effect in specific

Table 1 Specific heat jumps at 90 K (ΔC_p) and 220 K ($\Delta C_{p,h}$)

Sample	T_c (K)	ΔC_p ($\text{J mol}^{-1} \text{K}^{-1}$)	$\Delta C_{p,h}$ ($\text{J mol}^{-1} \text{K}^{-1}$)	$\Delta C_{p,h}/\Delta C_p$
UK1	90	1.1(2)	1.2(1)	1.1
UK3	91.5	1.67(5)	2.7(1)	1.6
B39/2	92.1	3.4(3)	2.4(1)	0.7
UK3NEG	—	0.4(2)	0.2(1)	0.6

heat we consider this possibility to be very unlikely. The observed jump in specific heat at the 90-K transition is in the range of earlier reported results², and as mentioned above, the anomaly at 220 K is largest in the samples with the largest jump at 90 K.

We acknowledge stimulating collaboration with K. A. Müller, J. G. Bednorz, C. E. Gough, J. S. Abell, M. A. Novotny, O. Trætteberg, T. Bye, F. Vassenden and T. Helgesen. This work has been supported by STATOIL (VISTA), the Royal Norwegian Council for Scientific and Industrial Research (NTNF), Norsk Hydro A/S and Elkem A/S.

Received 19 October; accepted 16 November 1987.

- Bednorz, J. G. & Müller, K. A. *Z. Phys.* **B64**, 189–193 (1986).
- Lagreed, T. *et al. Phys. Rev. B* (submitted).
- Cava, R. J. *et al. Nature* **329**, 423–425 (1987).
- Ourmazd, A. & Spence, J. C. H. *Nature* **329**, 425–427 (1987).
- Renault, A., Burdett, J. K. & Pouget J.-P. *J. Solid St. Chem.* (submitted).
- Bhargava, R. N., Herko, S. P. & Osborne, W. N. *Phys. Rev. Lett.* **59**, 1468–1471 (1987).
- Djurek, D. *et al. Phys. Lett. A* **123**, 481–484 (1987).
- Monod, P. *et al. J. Phys., Paris* **48**, 1369–1375 (1987).
- Stokka, S. & Fossheim, K. *J. Phys. E* **15**, 123–127 (1982).
- Kitazawa, K. *et al. Jap. J. appl. Phys.* **26**, L748–L750 (1987).
- Uher, C. & Kaiser, A. B. *Phys. Rev. B* (in the press).
- Jezowski, A. *et al. Phys. Lett. A* **122**, 431–433 (1987).
- Morelli, D. T., Heremans, J. & Swets, D. E. *Phys. Rev. B* **36**, 3917–3919 (1987).
- Lagreed, T. *et al. Phys. Scripta* (in the press).
- Capponi, J. J. *et al. Europhys. Lett.* **3**, 1301–1307 (1987).
- Simon, A. *Angew. Chem. int. Edn* **6**, 579–580 (1987).
- Segre, C. U. *et al. Nature* **329**, 227–229 (1987).
- Jorgensen, J. D. *et al. Phys. Rev. B* **36**, 3608–3616 (1987).
- Büttner, H. & Blumen, A. *Nature* **329**, 700–701 (1987).
- Phillips, J. C. *Phys. Rev. Lett.* **59**, 1856–1859 (1987).
- Cheong, S.-W. *et al. Phys. Rev. B* **36**, 3913–3916 (1987).
- Yusheng, H. *et al. J. Phys. F* **17**, L243–L248 (1987).
- Cannelli, C. *et al. Phys. Rev. B* (in the press).
- Jacobs, L., José, J. V., Novotny, M. A. & Goldman, A. *Europhys. Lett.* **3**, 1295–1300 (1987).
- Warren, W. W. Jr, Walstedt, R. E., Brennert, G. F., Espinosa, G. P. & Remeika, J. P. *Phys. Rev. Lett.* **59**, 1860–1863 (1987).

Radiation-related retrograde hydrogen isotope and K–Ar exchange in clay minerals

Ghislaine Halter*, Simon M. F. Sheppard†, Francis Weber‡, Norbert Clauer‡ & Maurice Pagel*

* Centre de Recherches sur la Géologie de l'Uranium, BP23, 54501 Vandoeuvre Cedex, France

† Centre de Recherches Pétrographiques et Géochimiques, BP20, 54501 Vandoeuvre Cedex, France

‡ Centre de Sédimentologie et de Géochimie de la Surface, 1 rue Blessig, 67084 Strasbourg Cedex, France

Hydrogen and oxygen isotope studies have been widely applied to characterize the origin of fluids during ore-forming processes^{1–4}. The primary isotope record, however, may be disturbed by retrograde exchange reactions, thus complicating the interpretation of the data. The susceptibility of minerals to retrograde isotope and chemical exchange is variable⁵, reflecting differences in the mechanism and rate of isotope exchange. These depend on several factors, including mineralogy, surface area, temperature, pressure, nature and concentration of fluid phase, water/rock ratio. In the

Cluff-D uranium deposit within the Athabasca basin (Saskatchewan, Canada) illites and chlorites associated with uranium mineralization are strongly depleted in deuterium ($-90 > \delta D > -170\%$ SMOW (standard mean ocean water)) and have young K/Ar apparent ages (<750 Myr) compared with illites from the barren zones around the ore deposit ($-48 > \delta D > -62\%$; $1,265$ Myr)⁶⁻⁸. Both the δD values and K/Ar 'ages' decrease with increase in the uranium content of the rock. The H_2O^+ content of illite from mineralized samples is higher than the theoretical value as well as samples from the barren zone. These relationships are interpreted in terms of radiation-catalysed retrograde hydrogen isotope and K-Ar exchange of the clays at low temperatures with post-Cretaceous meteoric waters.

The Athabasca basin (Saskatchewan, Canada) contains ten major high-grade uranium deposits ($\sim 300,000$ tons U, mean grade $\sim 2\%$ U). These deposits are spatially related to the major middle Proterozoic unconformity between the Aphebian-Archaeon basement and the Helikian cover⁹⁻¹¹. The newly formed minerals in the sediments are indicative of strong diagenetic¹² to anchimetamorphic conditions with temperatures of $\sim 200^\circ\text{C}$ (ref. 13). H- and O-isotope results^{7,14-16} indicate that waters in equilibrium with these minerals have δD ranging from -65 to -35% (SMOW) and $\delta^{18}\text{O}$ ranging from $+2$ to $+8\%$ (SMOW). Such isotope values and the salinities of the fluid inclusions¹³ are comparable to many formation waters from actual sedimentary basins¹⁷.

The Cluff-D deposit in the south of the Carswell circular structure in the Athabasca basin is located beneath the glacial cover. The D/H and $^{18}\text{O}/^{16}\text{O}$ ratios, H_2O^+ contents and K/Ar ages on mineral separates, whose purity was verified by X-ray diffraction, are given in Table 1 with the uranium content of the whole rocks. In the mineralized zone with $>1,000$ p.p.m. U, chlorites and illites are systematically depleted in deuterium by up to 120% relative to the same mineral either from the immediately surrounding alteration halo with <300 p.p.m. U or from the barren country rocks with <10 p.p.m. U (Fig. 1a). Illites and chlorites from the ore zone also have higher H_2O^+ contents and slightly higher $^{18}\text{O}/^{16}\text{O}$ ratios than those for the same mineral from outside of the ore zone (Table 1).

In the halo and mineralized zones, the illites can be termed hydromuscovites as they have high H_2O contents compared with muscovite and high K_2O contents. The increase in H_2O content up to 7.23 wt % in these illites as potassium decreases to 6.47 wt % (Table 1) probably reflects the substitution of K^+ by H_3O^+ (ref. 7). On approaching the mineralization in the Cluff-D deposit, the illite content decreases whilst the chlorite content increases¹⁷. The K_2O content of these chlorites increases with increase in the whole-rock U content and this cannot be related to the presence of illite in our samples.

A correlation is also observed between the apparent K/Ar age of the mineral and the uranium content of the rock (Fig. 1b), with the ages ranging from $1,320$ to 130 Myr. The Proterozoic K/Ar ages of $\sim 1,260$ Myr for the illites from the barren zones and the alteration halo are considered to be representative of the emplacement of the mineralization^{18,19}. The younger ages for the illites and chlorites do not appear to be simply related to either the decrease in K_2O content or the increase in H_2O content of the mineral. Similarly young and variable ages on illites have been reported from other Athabasca basin U deposits: K/Ar ages of 414 Myr at Midwest Lake¹⁶, $^{40}\text{Ar}/^{39}\text{Ar}$ plateau age of 639 Myr at McClean Lake¹⁹. Although 'normal' K/Ar and $^{40}\text{Ar}/^{39}\text{Ar}$ ages ($\sim 1,260$ Myr) have been reported on samples from alteration assemblages, including those spatially associated with mineralization^{16,19}, these results cannot be directly compared or contrasted with ours because the U content of the samples was not reported. The young ages are interpreted as apparent ages due to loss of Ar after mineralization^{7,16}. Thus, although these ages are geologically meaningless as ages, they emphasize that these systems have been disturbed to an extent dependent on their U content.

In the Cluff-D deposit, the U content of a sample is not controlled by the content of carbonaceous material²⁰, which also occurs in barren zones. This independence excludes a significant influence of deuterium-depleted Proterozoic organic matter ($-80 > \delta D > -350\%$ (ref. 21)) on the δD values of the clay minerals. The same fluid phase was therefore probably responsible for ore transport and alteration of the non-mineralized halo.

Table 1 Whole-rock uranium contents and mineral isotope and chemical data

Sample mineral*	U (p.p.m.)	$\delta D^\dagger$ (‰)	$\delta^{18}\text{O}^\dagger$ (‰)	H_2O^+ (wt %)	K_2O (wt %)	^{40}Ar (%)	^{40}Ar (10^{-6} cm ³ NTP/g)	K/Ar age (Myr)
Mineralization								
174 chlorite	—	-160	—	12.26	—	—	—	—
205 chlorite	4,000	-161	+6.6	12.13	1.20	50.36	5.22	130 ± 6
601 chlorite	4,100	-157	—	12.08	0.64	73.63	16.79	672 ± 22
638 chlorite	2,300	-139	—	10.83	0.50	70.47	15.97	432 ± 12
90.84 illite	2,300	-168	+12.1	7.23	6.47	89.90	191.47	742 ± 17
91.84 illite	2,420	-146	—	5.58	—	—	—	—
118.84 illite	1,050	-90	—	5.40	—	—	—	—
Halo alteration								
613 chlorite	165	-85	+5.8	9.50	0.37	78.02	15.71	989 ± 87
685 illite	253	-52	+9.7	5.17	8.36	99.80	525.08	$1,322 \pm 28$
686 illite	170	-51	—	5.36	8.79	99.84	522.63	$1,270 \pm 26$
89.84 illite	117	-79	—	4.82	—	—	—	—
153.84 illite	29	-61	—	5.17	—	—	—	—
310 illite	—	-51	+8.4	5.07	9.16	99.62	530.29	$1,245 \pm 26$
315 illite	72	-55	—	5.00	9.09	99.33	516.19	$1,228 \pm 26$
811 illite	17	-71	—	5.45	8.05	99.91	246.40	$1,292 \pm 29$
Barren								
832 illite	3	-52	—	4.93	7.58	99.79	448.03	$1,265 \pm 27$
154.84 illite	6	-62	+7.9	5.11	—	—	—	—
155.84 illite	4	-48	—	4.64	—	—	—	—
127.84 illite	4	-57	+12.2	4.62	—	—	—	—
121.84 kaolinite	5	-35	+13.4	13.93	—	—	—	—

* Less than $2 \mu\text{m}$ fraction; determined by X-ray diffraction.

† $\delta = (R_{\text{sample}}/R_{\text{standard}} - 1) \times 1,000$ where $R = \text{D/H}$ or $^{18}\text{O}/^{16}\text{O}$ and the standard is SMOW.

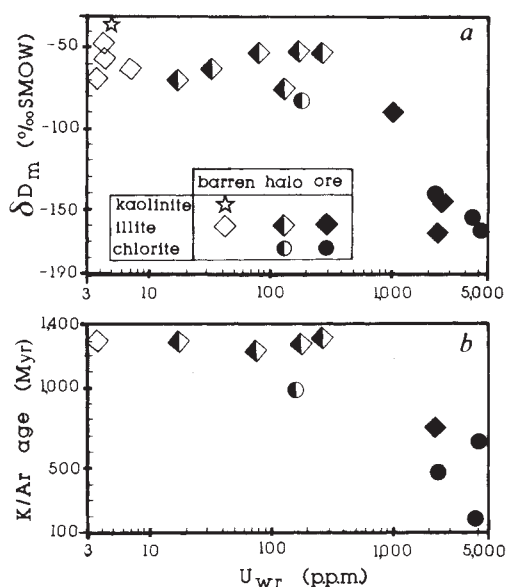


Fig. 1 Variation of the uranium content of the whole rock (wr) with: a, the hydrogen isotope composition of the mineral (δD_m value); and b, the K-Ar age of chlorite and illite.

All these results (Table 1 and Fig. 1) indicate that certain chemical and isotopic compositions of the clay minerals in the ore zone have undergone differential modification or retrograde exchange which is related to the uranium content of the rock. The conclusion that these phyllosilicates are Proterozoic minerals which were partially disturbed after their crystallization and not newly formed later minerals is also supported by the fact that: first, the index of crystallinity of illites, from within and outside of the ore zone are comparable; and second, the temperature of formation of 160–230 °C for the chlorites, estimated from the geothermometer of Cathelineau and Nieva²², is very similar to that determined for the formation of the gangue minerals by fluid inclusions (200 ± 50 °C)¹⁴. Field observations suggest that there is not a simple relationship between the degree of retrograde exchange and actual permeability (density of fractures) particularly as the highly mineralized facies are generally very clay rich and the adjacent alteration halo is as fractured as the ore zone. Nevertheless, most of the phyllites outside of the ore zone have not been significantly depleted in deuterium (Fig. 1a).

It is therefore proposed that the major radiation field associated with the high concentration of uranium has aided the retrograde exchange of the clay minerals in the ore zone. The interaction of alpha and beta particles and gamma rays with their environment causes ionizations, excitations, dissociations and/or displacements of atoms. Although more than three-quarters of the energy released by the radioactive decay of uranium is in the form of alpha particles, their distance of interaction in solids is limited to less than a few tens of micrometers²³. Our analysed clay minerals are not mixed with pitchblende on such a scale. The direct effect of alpha radiation can therefore be excluded as a general cause of the deuterium depletion process. In solids, beta particles and gamma rays can interact with atoms on a millimetre and centimetre scale respectively. The isotope exchange reactions must therefore be dominantly aided by the gamma-ray flux. The other radiations may contribute either if alpha emitters ²²⁶Ra and ²²²Rn migrate during the process, or through their 'excitation' of circulating waters which subsequently exchange with the illites or chlorites.

Isotope effects with three types of processes can be envisaged. (1) Kinetic isotope effect: this process would probably increase the δD value of the mineral as the deuterium bond is broken less frequently than the protium bond²⁴. The enrichment in $\delta^{13}C$

of carbonaceous material associated with high-grade U ore has been related to such an isotope effect induced by alpha irradiation²⁵. The reverse effect is observed in our samples. (2) Radiolysis: Dubessy *et al.*^{26,27} show that the occurrence of H₂ and O₂ under pressure in some aqueous fluid inclusions in mineralized sandstones at Cluff-D deposit is the result of the radiolysis of water. Assuming equilibrium-type fractionation factors, the H₂ is strongly depleted in deuterium relative to H₂O (~700‰ (ref. 28)). The remaining H₂O becomes enriched in deuterium to an extent depending on the degree of radiolysis. Exchange with clay mineral hydrogen should therefore increase their δD values. The reverse is observed. (3) Radiation-catalysed re-equilibration: the D/H ratio of the most retrogressively altered clay minerals could be approximately in isotope equilibrium with post-Cretaceous meteoric waters which have had low δD values of about -110‰ in the early Tertiary¹ to present day values of about -150‰ (ref. 29). It is proposed that the extent of hydrogen isotope exchange with low-temperature post-Cretaceous meteoric waters and loss of argon are related to the intensity of the radiation field generated from the associated uranium. The radiation field acts as a catalyst by making the mineral more fragile and therefore more susceptible to isotopic and chemical exchange. Equally, ionization of water near the uranium may also aid the clay-H₂O exchange reactions.

The radiation-catalysed re-equilibration hypothesis does not imply that clay minerals, which are not directly associated with high uranium contents (>1,000 p.p.m. U), cannot undergo retrograde exchange. Some clay minerals which are on fractures have undergone exchange or recrystallization¹⁶. However, the great majority of clays outside of the ore zone seem to have retained their initial compositions to a very good approximation. This is also in agreement with isotope systematics on most clay minerals of different ages from a wide variety of environments so long as no significant mineralogical changes have occurred^{30–34}. The increase in H₂O⁺ content of illite and chlorite in the ore zone attests that subtle changes in the minerals occurred as a consequence of the high radiation field.

We thank Amok Ltd and their geologists for their support, J. Dubessy and B. Poty for discussions, P. Coget, R. A. Wendling and R. Winkler for technical assistance, and A. E. Fallick and J. Hunziker for their reviews. This research was supported by ATP Gîtologie, PIRSEM-CNRS.

Received 28 July; accepted 15 October 1987.

1. Taylor, H. P. Jr *Econ. Geol.* **69**, 843–883 (1974).
2. Ohmoto, H. *Rev. Miner.* **16**, 491–559 (1986).
3. Sheppard, S. M. F. *Rev. Miner.* **16**, 165–183 (1986).
4. Taylor, B. E. in *Stable Isotope Geochemistry of Low Temperature Processes* (ed. Kyser, T. K.) 337–445 (Miner. Ass. Canada, 1987).
5. Cole, D. R. & Ohmoto, H. *Rev. Miner.* **16**, 41–90 (1986).
6. Halter, G., Sheppard, S. M. F., Pagel, M. & Weber, F. *Terra Cognita* **5**, 153 (1985).
7. Halter, G., Pagel, M., Sheppard, S. M. F., Weber, F. & Clauer, N. *Mémoires du BRGM* (in the press).
8. Clauer, N., Ey, F. & Gauthier-Lafaye, F. in *The Carswell Structure Uranium Deposits, Saskatchewan* (eds Laine, R., Alonso, D. & Svab, M.) 47–53 (Spec. Pap. Geol. Ass. Canada, 29, 1985).
9. Pagel, M. & Jaffrezic, H. *C. R. hebdom. Acad. Séanc. Acad. Sci., Paris* **284D**, 113–116 (1977).
10. Hoeve, J. & Sibbald, T. I. I. in *Symp. Uranium Saskatchewan* (ed. Dunn, C. E.) 331–354 (Saskat. Geol. Surv., spec. publ. no. 3, 1976).
11. Ey, F. thesis, Univ. Strasbourg (1984).
12. Hoeve, J., Rawsthorn, K. & Quirt, D. *Saskat. Geol. Surv. Misc. Rep.* **81–84**, 76–89 (1981).
13. Pagel, M. *C. R. hebdom. Séanc. Acad. Sci., Paris* **284D**, 113–116 (1975).
14. Pagel, M., Poty, B. & Sheppard, S. M. F. in *Uranium in the Pine Creek Geosyncline* 639–654 (IAEA, Vienna, 1980).
15. Wallis, R. H., Saracoglu, N., Brummer, J. J. & Golightly, J. P. *CIM Bull.* 1–28 (1984).
16. Wilson, M. R., Kyser, T. K., Mehnert, H. H. & Hoeve, J. *Geochim. cosmochim. Acta* **51**, 869–878 (1987).
17. Clayton, R. N., Friedman, I., Graf, D. L., Mayeda, T. K. & Shimp, N. F. *J. geophys. Res.* **71**, 3869–3882 (1966).
18. Ey, F., Gauthier-Lafaye, F., Lillie, F., Schumacher, F. & Weber, F. in *The Carswell Structure Uranium Deposits, Saskatchewan* (eds Laine, R., Alonso, D. & Svab, M.) 121–138 (Spec. Pap. Geol. Ass. Canada, 29, 1985).
19. Bray, C. J. *et al. Can. J. Earth Sci.* **24**, 10–23 (1987).
20. Ruhlmann, F. in *The Carswell Structure Uranium Deposits, Saskatchewan* (eds Laine, R., Alonso, D. & Svab, M.) 105–120 (Geol. Ass. Canada, spec. paper 29, 105–120, 1985).
21. Schoell, M. *Org. Geochem.* **6**, 645–663 (1984).
22. Cathelineau, M. & Nieva, D. *Contrib. miner. Petrol.* **91**, 235–244 (1985).
23. Faul, H. (ed.) *Nuclear Geology* 1–414 (Wiley, New York, 1954).
24. Melander, L. *Isotope Effects on Reaction Rates* 1–181 (Ronald, New York, 1960).
25. Leventhal, J. S. & Threlkeld, C. N. *Science* **202**, 430–432 (1978).

26. Dubessy, J., Beny, J. M., Pagel, M., Poty, B. & Kosztolanyi, C. *IMA Meeting, Orléans, Collected Abstracts* 129 (1980).
27. Dubessy, J. *et al. Geochim. cosmochim. Acta* (submitted).
28. Friedman, I. & O'Neil, J. R. in *Data of Geochemistry* 6th edn, 440-KK (US Govt Printing Office, Washington DC, 1977).
29. Friedman, I., Redfield, A. C., Schoen, B. & Harris, J. *Rev. Geophys.* **2**, 177-224 (1964).
30. Savin, S. & Epstein, S. *Geochim. cosmochim. Acta* **34**, 25-42 (1970).
31. Sheppard, S. M. F., Nielsen, R. L. & Taylor, H. P. Jr *Econ. Geol.* **64**, 755-777 (1969).
32. Lawrence, J. R. & Taylor, H. P. Jr *Geochim. cosmochim. Acta* **36**, 1377-1393 (1972).
33. Sheppard, S. M. F. *J. geol. Soc.* **133**, 573-591 (1977).
34. Clauer, N., O'Neil, J. R. & Bonnot-Courtois, C. *Geochim. cosmochim. Acta* **46**, 1755-1762 (1982).

Restricted utility of the pristane/phytane ratio as a palaeoenvironmental indicator

H. L. ten Haven*, J. W. de Leeuw†, J. Rullkötter* & J. S. Sinninghe Damsté†

* Institute of Petroleum and Organic Geochemistry (ICH-5), KFA Jülich, PO Box 1913, D-5170 Jülich, FRG

† Organic Geochemistry Unit, Delft University of Technology, Department of Chemistry and Chemical Engineering, de Vries van Heystplantsoen 2, NL-2628 RZ Delft, The Netherlands

The acyclic C_{19} and C_{20} isoprenoid hydrocarbons, pristane (Pr) and phytane (Ph), respectively, have been widely assumed to be diagenetic products of the phytol side chain of chlorophyll¹⁻³, although alternative sources of precursors have been suggested. The ratio of these two compounds is usually interpreted to be an indicator of the oxicity of the environment of deposition^{3,4}. Recent advances in organic geochemistry in combination with geological constraints lead us to suggest that the Pr/Ph ratio cannot be used as an indicator for oxygen levels. However, in hypersaline environments of deposition the rationale behind a low Pr/Ph ratio is easier to understand, and in these environments application of the Pr/Ph ratio can be expected to be successful.

As early as 1969 Brooks *et al.*² suggested that variations in the Pr/Ph ratio may reflect variations in the degree of oxidation during early diagenesis of chlorophyll. In extension of this, Didyk *et al.*⁴ proposed a direct relationship of the Pr/Ph ratio with the oxicity of the environment of deposition, but they also emphasized that other lines of evidence, such as chlorin contents, should be considered in order to assess properly the depositional environment. A Pr/Ph ratio of less than unity, according to these authors, would indicate an anoxic and thus reducing environment, and *vice versa* for a ratio greater than unity. With strict adherence to these guidelines, but without paying attention to the limitations, this invoked application of the ratio as an environmental indicator has had an enormous impact on the organic geochemical literature, because both isoprenoid compounds are almost ubiquitously present in the geosphere. Many organic geochemists and petroleum geochemists use this parameter, partly because of its easy measurement by simple gas chromatography of the hydrocarbon fraction of sediment extracts or crude oils.

The rationale of the Pr/Ph ratio as an environmental indicator is that in an anoxic environment reducing diagenetic pathways prevail over oxidative pathways. The isoprenoid alcohol, phytol, assumed to be liberated from chlorophyll at an early stage of diagenesis by hydrolysis, is taken as starting point for a series of defunctionalization reactions¹⁻⁴. While under reducing conditions phytol may ultimately be transformed to phytane, it is hypothesized that under oxic conditions pristane is generated from phytol via decarboxylation of an intermediate carboxylic acid.

The recent suggestion that archaeobacterial lipids (for example, from methanogens or halophiles) could be another source for these isoprenoids^{5,6} in combination with a possible origin of pristane from tocopherols⁷, has greatly weakened the rationale for the use of the Pr/Ph ratio. In addition, pristane and pristanes

have also been detected in zooplankton^{8,9}, although such an origin of sedimentary pristane has never really been considered to be a major source. Furthermore, an analytical problem of precise determination of Pr/Ph ratios arises from the recent identification of an unusually branched alkane (2, 6, 10-trimethyl-7-(3-methylbutyl)-dodecane) in sediments, which coelutes with pristane on most capillary columns¹⁰; this raises doubt about the validity of Pr/Ph ratios previously published in the literature.

Geological constraints are rarely considered. Invariably ignored is the geological fact that all fine-grained nearshore, shelf and hemipelagic sediments are anoxic below a surface oxic layer, the thickness of which depends on several factors, the most important one being the rate of sedimentation¹¹. Step-wise extraction techniques, applied to decipher the mode of occurrence of lipids in organic-matter-rich sediments from the eastern Mediterranean, covering a time span from 1.10³ to 2.10⁵ years BP, revealed that most of the phytol encountered (>95%) was present in an esterified form and in abundant quantities^{12,13}, while alleged phytol transformation products (for example, phytadienes, dihydrophytol, phytanic acid, phytane, pristane⁴) were absent or present in very low concentrations. When phytol is eventually degraded in these sediments, this process will take place under reducing conditions, giving rise to low Pr/Ph ratios following the scheme of Didyk *et al.*⁴. Hence, this ratio would not be expected to reflect the oxicity of the environment of deposition, but would be merely the result of reactions taking place well after reducing conditions have been established.

Gravity-driven mass transport is another geological phenomenon which can lead to a misinterpretation of the Pr/Ph ratio. Turbidites, enriched in organic matter, have been found interbedded in pelagic sediments at several locations^{14,15}. Progressive subsurface oxidation fronts greatly alter the labile organic matter in these turbidites¹⁴. Assuming that most of the phytol is still esterified to chlorophyll, such a post-depositional oxidation effect would result in elevated Pr/Ph ratios, which, however, do not reflect the original environment of deposition of these organic-matter-rich sediments. It has also been suggested that microenvironments, which do not reflect the redox conditions of the sediment as a whole, may be active sites where pristane and phytane are formed via microbially mediated reactions⁶.

From the above discussion it is clear that it is virtually impossible to draw valid conclusions from the Pr/Ph ratio with respect to the oxicity of the environment of deposition, although this does not imply that we refute the observation that sediments deposited under strictly anoxic conditions are often characterized by low Pr/Ph ratios⁴. An alternative explanation for this observation is that pristane reflects the primary production of organic matter (represented by chlorophylls, tocopherols) and that most of the phytane stems from the strictly anaerobic methanogenic bacteria^{5,6}, which after burial flourish on organic matter well preserved as a result of anoxic conditions during deposition. However, in this respect it should be noted that phytanyl ether lipids are of minor importance in methanogens in comparison with biphytanyl ether lipids¹⁶ which should not produce phytane very easily during diagenesis.

A special case, in which there may be some application of the Pr/Ph ratio as an environmental indicator, seems to be a hypersaline environment of deposition^{17,18}. Reported Pr/Ph ratios of hypersaline environments are presented in Table 1. Low values are clearly characteristic, but this criterion alone is, of course, not sufficient to identify such an environment positively. An example of such an exception are the Pr/Ph ratios of the sediments from the Paradox Basin (Table 1). Other biological marker characteristics should therefore be considered¹⁷⁻²⁰, if possible, supported by geological information.

Hypersaline environments can be zones of high biological productivity, but with a low diversity of organisms^{21,22}; of these

Table 1 Pristane/phytane ratios of hypersaline environments

Sample	Location	Age	Pr/Ph	Ref.
Mediterranean area				
Rocks (3)	Northern Apennines	Miocene	0.1	18
Oils (3)	Sicily	Miocene	0.01–0.1	17, 27
Oil	Prinos, Greece	?	0.3	28
Oils (5)	East Mediterranean	Tertiary	0.3–0.8	29
Oil	Greece	Miocene	0.25	30
Rocks (20)	East Mediterranean	Tertiary	0.4–0.8	29
Oil	Amposta, Spain	Miocene	0.61	31
Rock	Israel	Cretaceous	0.45	32, 33
Rock	Jordan	Cretaceous	0.47*	34
Central Europe				
Rocks (3)	Rhine graben	Lower Miocene/Upper Oligocene	0.05–0.14	35
Rocks (17)	Rhine graben	Lower Miocene/Upper Oligocene	0.02–0.7	29
Oils (3)	Rhine graben	Lower Miocene/Upper Oligocene	0.05–0.4	29
Rocks (10)	South Germany	Triassic	0.36–0.53	36
Rock†	Aquitaine, France	Jurassic	1.20	37
Rock‡	Aquitaine, France	Jurassic	0.30	37
Oils(?)	Aquitaine, France	?	0.48–0.90	37
Rock	Scotland	Devonian	0.6	38
Rocks (28)	Germany	Permian + Miocene	<1	39
China				
Rocks (12)	Jiangnan Basin	Oligocene	0.05–0.46	40
Oils (2)	Jiangnan Basin	Oligocene	0.08–0.18	41
Oils (2)	Chaidamu Basin	Lower Miocene/Upper Oligocene	0.27–0.30	41
North America				
Oils (33)	Michigan Basin	Silurian	0.4–0.5	42
Oils (3)	Michigan Basin	Silurian	~0.56	43
Oils (8)	Michigan Basin	Silurian	~0.47	44
Oils (18)	Ontario	Silurian	0.47–0.66	45
Oil	Alberta	Devonian	0.15	46
Rocks (19)	Williston Basin	Triassic	~0.70	47
Oils (7)	South Florida	Cretaceous	0.54–0.73	48, 49
Rocks (7)	Paradox Basin	Carboniferous	1.1–1.4	50
Oil	Rozel Point, Utah	Miocene	0.1	17
Other areas				
Rocks (9)	Guatemala	Cretaceous	0.32–0.75	51
Oils (?)	Guatemala	Cretaceous	~0.6	37, 48
Oils (10)	Australia	Cambrian	0.82–1.10	52

The reported Pr/Ph ratios are maximum values, as a coelution of 2,6,10-trimethyl-7-(3-methylbutyl)-dodecane with pristane would decrease the pristane/phytane ratio. Numbers in parentheses give number of samples tested.

* Sample reanalysed in Delft Laboratory. A hypersaline environment is inferred from characteristic biological markers^{17,18} and literature descriptions from adjacent areas of the same age³³.

† Sample with 30% anhydrite and 61% dolomite.

‡ Sample with 70% anhydrite and 28% dolomite.

halophilic bacteria are very important. These bacteria are known to contain complex lipids with a phytanyl moiety²³ in much higher amounts than in methanogenic bacteria, which can explain why sediments from hypersaline environments are characterized by low Pr/Ph ratios. At a certain level of salinity halophilic bacteria will have their prolific growth stage and it is, therefore, postulated that the salinity will be the key control for the Pr/Ph ratio in such environments.

An increase of the Pr/Ph ratio with depth has been observed in uniform sedimentary sequences^{24,25} as well as in non-uniform strata^{2,26}. Apparently, maturation has an influence on the Pr/Ph ratio, although this cannot account for the results obtained for the relatively immature sediments from the Paradox Basin. Immature oils and sediment extracts listed in Table 1 are, where reported, characterized by high Ph/*n*-octadecane ratios (>1). Due to the generation and release of hydrocarbons during maturation the Ph/*n*-octadecane ratios will decrease and will eventually reach values below unity, while at the same time the Pr/Ph ratio will increase. Again it must be stressed that the above outlined scheme cannot be quantified as it is probably highly dependent on the origin of the organic matter.

This work was partly supported by the Netherlands Founda-

tion for Earth Science Research (AWON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO) and by the Deutsche Forschungsgemeinschaft (DFG).

Received 28 July; accepted 15 October 1987.

- Bendoraitis, J. G., Brown, B. L. & Hapner, L. S. *Analyt. Chem.* **34**, 49–53 (1962).
- Brooks, J. D., Gould, K. & Smith, J. W. *Nature* **222**, 257–259 (1969).
- Powell, T. G. & McKirdy, D. M. *Nature* **243**, 37–39 (1973).
- Didyk, B. M., Simoneit, B. R. T., Brassell, S. C. & Eglinton, G. *Nature* **272**, 216–222 (1978).
- Risatti, J. B., Rowland, S. J., Yon, D. A. & Maxwell, J. R. in *Advances in Organic Geochemistry 1983* (eds Schenck, P. A., de Leeuw, J. W. & Lijmbach, G. W. M.) 93–104 (Pergamon, Oxford, 1984).
- Volkman, J. K. & Maxwell, J. R. in *Biological Markers in the Sedimentary Record* (ed. Johns, R. B.) 1–42 (Elsevier, Amsterdam, 1986).
- Goossens, H., de Leeuw, J. W., Schenck, P. A. & Brassell, S. C. *Nature* **312**, 440–442 (1984).
- Blumer, M., Mullin, M. M. & Thomas, D. W. *Science* **140**, 974 (1963).
- Blumer, M. & Snyder, W. D. *Science* **150**, 1588–1589 (1965).
- Yon, D. A., Maxwell, J. R. & Ryback, G. *Tetrahedron Lett.* **23**, 2143–2146 (1982).
- Calvert, S. E. in *Marine Petroleum Source Rocks* (eds Brooks, J. & Fleet, A. J.) 137–151 (Blackwell, Oxford, 1987).
- ten Haven, H. L., Baas, M., de Leeuw, J. W. & Schenck, P. A. *Mar. Geol.* **75**, 137–156 (1987).
- ten Haven, H. L., Baas, M., de Leeuw, J. W., Schenck, P. A. & Brinkhuis, H. *Chem. Geol.* **64**, 149–167 (1987).
- Colley, S., Thomson, J., Wilson, T. R. S. & Higgs, N. C. *Geochim. cosmochim. Acta* **48**, 1223–1235 (1984).
- Wilson, T. R. S., Thomson, J., Colley, S., Hydes, D. J. & Higgs, N. C. *Geochim. cosmochim. Acta* **49**, 811–822 (1985).

16. Tornabene, T. G. & Langworthy, T. A. *Science* **203**, 51–53 (1979).
17. ten Haven, H. L. *et al.* in *Lacustrine Petroleum Source Rocks* (eds Kelts, K., Fleet, A. J. & Talbot, M.) (Blackwell, Oxford, in the press).
18. ten Haven, H. L., de Leeuw, J. W. & Schenck, P. A. *Geochim. cosmochim. Acta* **49**, 2181–2191 (1985).
19. Connan, J., Hussler, G. & Albrecht, P. in *AAPG Stud. Geol.* **18**, 206 (1984).
20. Powell, T. G. *Mar. Petr. Geol.* **3**, 200–219 (1986).
21. Kirkland, D. W. & Evans, R. *Bull. Am. Ass. Petrol. Geol.* **65**, 184–190 (1981).
22. Warren, J. K. *J. Sed. Petr.* **56**, 442–454 (1986).
23. Kates, M., Palameta, B., Joo, C. N., Kushner, D. J. & Gibbons, M. E. *Biochemistry* **5**, 4092–4099 (1966).
24. Tissot, B., Debyser, Y. C., Deroot, G. & Oudin, J. L. *Bull. Am. Ass. Petrol. Geol.* **55**, 2177–2193 (1971).
25. Albrecht, P., Vandenbroucke, M. & Mandengue, M. *Geochim. cosmochim. Acta* **40**, 1241–1249 (1976).
26. Rashid, M. A. *Chem. Geol.* **25**, 109–122 (1979).
27. Palmer, S. E. & Zumbeke, J. E. in *Organic Maturation Studies and Fossil Fuel Exploration* (ed. Brooks, J.) 393–426 (Academic, London, 1981).
28. Moldovan, J. M., Seifert, W. K. & Gallegos, E. J. *Bull. Am. Ass. Petrol. Geol.* **69**, 1255–1268 (1985).
29. Hollerbach, A. *Erdöl, Kohle, Erdgas, Petrochem.* **32**, 327 (1980).
30. Tissot, B. P. & Welte, D. H. *Petroleum Formation and Occurrence* (Springer, Berlin, 1978).
31. Albaiges, J., Algaba, J., Clavell, E. & Grimalt, J. in *Advances in Organic Geochemistry 1985* (eds Leythaeuser, D. & Rullkötter, J.) 441–450 (Pergamon, Oxford, 1986).
32. Spiro, B. in *Advances in Organic Geochemistry 1983* (eds Schenck, P. A., de Leeuw, J. W. & Lijmbach, G. W. M.) 543–559 (Pergamon, Oxford, 1984).
33. Spiro, B. & Aizenshtat, Z. *Nature* **269**, 235–237 (1977).
34. Wehner, H. & Hufnagel, H. *Mitt. Geol. Paläont. Inst. Univ. Hamburg* **60**, 381–395 (1986).
35. Waples, D. W., Haug, P. & Welte, D. H. *Geochim. cosmochim. Acta* **38**, 381–387 (1974).
36. Leythaeuser, D., Schaefer, R. G. & Radke, M. *Geologica bavarica* **81**, 159–192 (1981).
37. Connan, J. in *Petroleum Geology in China* (ed. Mason, J. F.) 48–70 (PennWell, Tulsa, 1981).
38. Hall, P. B. & Douglas, A. G. in *Advances in Organic Geochemistry 1981* (eds Bjorøy, M. *et al.*) 576–587 (Wiley, Chichester, 1983).
39. Welte, D. H. & Waples, D. W. *Naturwissenschaften* **60**, 516–517 (1973).
40. Fu, J., Sheng, G., Peng, P., Brassell, S. C., Eglinton, G. & Jiang, J. in *Advances in Organic Geochemistry 1985* (eds Leythaeuser, D. & Rullkötter, J.) 119–126 (Pergamon, Oxford, 1986).
41. Wang, T., Fan, P. & Swain, F. M. in *Lacustrine Petroleum Source Rocks* (eds Kelts, K., Fleet, A. J. & Talbot, M.) (Blackwell, Oxford, in the press).
42. Illich, H. A. *Bull. Am. Ass. Petrol. Geol.* **67**, 385–393 (1983).
43. Vogler, E. A., Meyers, P. A. & Moore, W. A. *Geochim. cosmochim. Acta* **45**, 2287–2293 (1981).
44. Pruitt, J. W. *Geochim. cosmochim. Acta* **47**, 1157–1159 (1983).
45. Powell, T. G., MacQueen, R. W., Barker, J. F. & Bree, W. G. *Bull. Can. Petrol. Geol.* **32**, 289–312 (1984).
46. Milner, C. W. D., Rogers, M. A. & Evans, C. R. *J. geochem. Explor.* **7**, 101–153 (1977).
47. Arnett, J. D. *Chem. Geol.* **46**, 113–140 (1984).
48. Palacas, J. G. *Proc. 11th World Petrol. Congr.* **2**, 31–43 (1984).
49. Palacas, J. G., Anders, D. E. & King, J. D. in *AAPG Stud. Geol.* **18**, 71–96 (1984).
50. Hite, R. J., Anders, D. E., Ging, T. G. in *Hydrocarbon Source Rocks of the Greater Rocky Mountains Region* (eds Woodward, J., Meissner, F. F. & Clayton, J. L.) 255–274 (Rocky Mountain Association of Geologists, Denver, 1984).
51. Connan, J., Bourouillec, J., Dessort, D. & Albrecht, P. in *Advances in Organic Geochemistry 1985* (eds Leythaeuser, D. & Rullkötter, J.) 29–50 (Pergamon, Oxford, 1986).
52. McKirdy, D. M., Kantsler, A., Emmett, J. K. & Aldridge, A. K. in *AAPG Stud. Geol.* **18**, 13–31 (1984).

Factors controlling emission of dimethylsulphide from salt marshes

John W. H. Dacey*, Gary M. King†
& Stuart G. Wakeham*‡

* Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

† I. C. Darling Center, University of Maine, Walpole, Maine 04573, USA

The emission of biogenic sulphur gases constitutes about half the atmospheric budget for gaseous sulphur¹. Since dimethylsulphide (DMS) was first implicated as a major component of gaseous sulphur flux^{2–4}, considerable attention has been given to its emission from various ecosystems. Salt marshes have been identified as one system with a high area-specific sulphur emission^{5–13}. Dimethylsulphide and hydrogen sulphide (H₂S) constitute the bulk of the flux from salt marshes, with DMS predominating in vegetated areas of the marsh^{6,8–13}. As H₂S is a product of anaerobic decomposition in sediments, it has been assumed that other sulphur gases emitted from salt marshes also originate from decomposition in sediment processes⁵. Our research suggests an alternative explanation for DMS fluxes. We have investigated the distribution of DMS and dimethylsulphoniopropionate (DMSP) in salt marshes

and conclude that DMS arises primarily from physiological processes in leaves of higher plants, mainly one species of grass, *Spartina alterniflora*. Furthermore, the emission of DMS from this grass may be influenced by the technique used to measure emission, and emission from sites dominated by *S. alterniflora* cannot be considered to be representative of marsh flora.

If we assume that DMS and H₂S emission from salt marshes result from processes in the top 2 cm of sediment where plant root biomass and microbial activity are typically greatest¹⁴, it is possible to estimate turnover rates for the dissolved gas pools. The observed flux of H₂S from salt-marsh sediments (150 $\mu\text{mol m}^{-2} \text{d}^{-1}$)⁹ represents a turnover of only 0.1% per day for typical porewater concentrations of 500 $\mu\text{mol l}^{-1}$ (refs 14, 15). In contrast, DMS emission from sediments with undisturbed plants (1–246 $\mu\text{mol m}^{-2} \text{d}^{-1}$)⁹ would require turnover rates of 100–30,000% per day for typical porewater concentrations of 50 nmol l^{-1} (ref. 15). This discrepancy suggests that the sources and controlling factors of the two major sulphur gas emissions from salt marshes must differ markedly.

We have investigated the concentration of DMSP in a variety of higher plants inhabiting the intertidal salt marshes of eastern North America. The pool of DMSP is especially high in the tissues of *S. alterniflora* (Table 1). We have not recorded DMSP

Table 1 Contents of leaves of major American salt-marsh species

Marsh species	Location	DMSP concentration ($\mu\text{mol per g dry weight}$)
<i>Spartina alterniflora</i>	MA, DE, SC	80–280
<i>Spartina patens</i>	MA, DE	<0.1
<i>Spartina cynosuroides</i>	DE	<0.1
<i>Distichlis spicata</i>	MA, DE	<0.1
<i>Juncus gerardi</i>	MA	<0.1
<i>Juncus roemerianus</i>	SC	<0.1
<i>Salicornia europaea</i>	MA	<0.1
<i>Phragmites communis</i>	MA	<0.1
<i>Rhizophora mangle</i>	VI	<0.1

Samples were taken during the growing season. Seasonal and geographical variations were minimized by always collecting specimens of *S. alterniflora* with the other plant samples (except in the case of *R. mangle*). The European marsh grass *S. anglica* is reported to contain up to 200 $\mu\text{mol DMSP g}^{-1}$ dry weight²⁰. KOH digestion of 13 MeOH extracts of leaves of *S. alterniflora* produced an acrylic acid to DMS ratio of 1.04 (± 0.04 , s.e.m.), indicating that all the DMS liberated during base digestion arose from DMSP. Rapid extraction of plant tissues in the field precluded production of acrylate and loss of DMS as an artefact. This elimination reaction may occur during the normal course of plant metabolism or as a function of cell death or turnover.

MA, Massachusetts; DE, Delaware; SC, South Carolina; VI, Virgin Islands.

at concentrations greater than 0.1 $\mu\text{mol per g dry weight}$ in tissues of any other marsh species. Perhaps coincidentally, there are no published reports of measurements of DMS emission from any other marsh species or from marshes lacking *S. alterniflora*. Previously, DMSP has been documented in a wide range of marine algae^{16–18} and in two higher plants, *Spartina anglica*^{19,20} and *Zostera marina*¹⁸. In certain species, DMSP seems to be involved in regulating cellular osmotic pressure²¹ and may also be important in sulphur storage²⁰. DMSP is thought to degrade mainly by an enzymatically catalysed elimination reaction, yielding DMS and acrylic acid¹⁶. We propose that this reaction is carried out in the leaves of *S. alterniflora* and is the major source of the observed flux of DMS from salt marshes. Similar processes should dominate wherever vegetation contains high concentrations of DMSP: in *S. anglica*²⁰ and in many intertidal algal genera such as *Ulva*¹¹. Degradation of

‡ Present Address: Skidaway Institute of Oceanography, Savannah, Georgia 31416, USA.

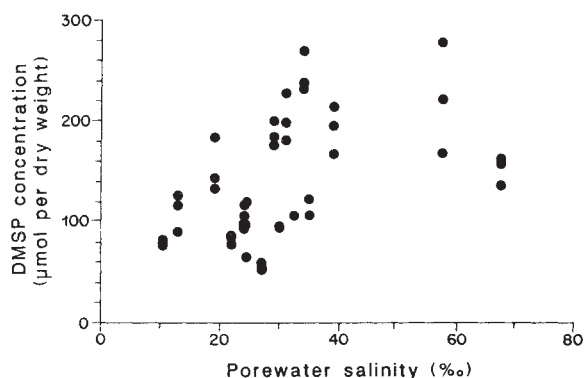


Fig. 1 Concentration of DMSP in leaves of *S. alterniflora* as a function of sediment salinity (0–2 cm depth). Samples were collected from marshes in Massachusetts and Delaware at various times during the year. DMSP content in leaves is influenced by other environmental factors (see Table 2).

DMSP and other sulphonium compounds (for example S-methylmethionine) in higher plants may also be important in terrestrial systems, where DMS seems to be a major component of sulphur flux to the atmosphere^{22,23}

Concentrations of DMSP in *S. alterniflora* were found to be highest in leaves (80–300 $\mu\text{mol per g dry weight}$) and lowest in roots and rhizomes (20–60 $\mu\text{mol per g dry weight}$). Based on conservative estimates of leaf and root concentrations of 120 and 40 $\mu\text{mol per g dry weight}$ respectively, the total amount of DMSP integrates to $\sim 60 \text{ mmol DMSP m}^{-2}$ in the Great Sippewissett Marsh, Falmouth, Massachusetts. On the other hand, there is only $\sim 2 \mu\text{mol DMS m}^{-2}$ of salt-marsh surface in porewater from the 0–2 cm depth interval. The fact that DMSP is at least 30,000 times more abundant than DMS, and that DMSP degrades physiologically to DMS¹⁶, suggests that a very small turnover of plant DMSP could be responsible for the observed DMS flux. A turnover of only 0.4% of this DMSP per day would generate sufficient DMS to support the highest DMS emission rates measured in salt marshes. Such a low rate of DMSP turnover is similar to previously observed DMS production rates in marine algae where $\sim 1\%$ of DMSP is converted to DMS per day^{24,25}.

The concentration of DMSP in leaves of *S. alterniflora* generally increases with increasing sediment salinity (Fig. 1) and our data indicate that the concentration of DMSP in *S. alterniflora* is influenced by the nutritional status of the grass. Plants fertilized with nitrogen contain less DMSP and more glycinebetaine (which is a nitrogen-bearing osmotic effector in marsh grass) in their tissues (Table 2).

To evaluate the role of plant species on DMS emission from salt marshes, we collected individual cores of sediment-bearing *S. alterniflora*, *S. patens* and *Distichlis spicata*. Substantial accumulation of DMS occurred in the headspace over *S. alterni-*

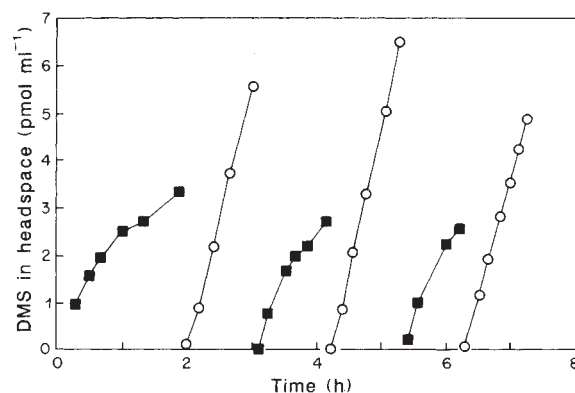


Fig. 2 Accumulation of DMS in the headspace over *S. alterniflora*. At roughly hourly intervals, the headspace was opened to the atmosphere and flushed out. In consecutive runs, water was either added or removed from the surface of the sediment: Circles, surface flooded; squares, surface exposed. DMS accumulated more rapidly when the sediment was flooded with 2 cm water. The concentration of DMS in porewater at $\sim 1 \text{ cm}$ depth was determined by withdrawing water through the side of the core by syringe. DMS concentration in porewater was 50 nmol l^{-1} , and in surface water, 35 nmol l^{-1} , corresponding to about 4 pmol ml^{-1} and 3 pmol ml^{-1} , respectively, in the headspace at equilibrium. The flux represented by the linear increase in headspace concentration with flooded sediment is $2.2 \mu\text{mol m}^{-2} \text{ h}^{-1}$. This flux represents a turnover of about 0.8% of leaf DMSP per day versus a turnover of 520% per day for porewater DMS in the entire 20 cm of sediment in the core.

flora, while emission rates from the other two species were more than 100 times smaller.

Dimethylsulphide in the headspace of the *S. alterniflora* core was emitted by the leaves. When the surface of the sediment in the core was covered with 2 cm of sea water, rates of DMS accumulation in the headspace of the core were substantially higher than when the sediment was exposed (Fig. 2). The concentration of DMS in the headspace reached levels three times higher than equilibrium with surface water and two times higher than with sediment porewater (Fig. 2). This indicates that the DMS could not have diffused or bubbled out of the sediment; it must have entered the headspace from the leaves. The data in Fig. 2 suggest that the sediment acted as a sink for DMS rather than a source. Covering the sediment with water slowed the flux of DMS from the headspace to the sediment, thereby lowering its rate of consumption.

These data suggest that more attention must be directed towards the method of gas emission measurement. Clearly the factors controlling DMS and H_2S emissions are distinct. Therefore, different procedures may be required to measure the rate of emission of each gas. Most of the H_2S emitted from salt marshes probably passes directly across the sediment surface, although some small amount may be produced by plants²⁶. In contrast, the emission of DMS requires that the physiology of DMSP in the marsh grasses be considered. As DMSP seems to play an important part in osmotic adjustment in these plants, the dynamics of DMSP in leaves may be influenced by short-term changes in the heat and water balance of leaves. The common practice of enclosing marsh plants in a chamber may lead to erroneous estimates of DMS emission, even in flow-through systems where flow rate is insufficient to minimize perturbations of leaf temperature and gas exchange (H_2O and CO_2). Flow-through systems that strip sulphur gases from the inlet stream^{6,12} may also lower CO_2 partial pressures in the chamber, thus altering physiological processes and therefore emission of DMS. Commercial cylinder air may be a poor choice for sweep gas in chambers¹¹ as it frequently contains very little CO_2 ($< 3 \text{ p.p.m.}$, compared with $> 330 \text{ p.p.m.}$ in air) which must affect leaf physiology. Clipping of leaves in chambers⁷ must also influence DMS emission.

Table 2 DMSP and glycinebetaine (GBT) content of fertilized and control *Spartina alterniflora*

Plant type	Concentration in leaves ($\mu\text{mol per g dry weight}$)	
	DMSP	GBT
Short grass	136 (28)	76 (3)
Fertilized short grass	89 (15)	146 (23)
Tall creekbank grass	84 (5)	94 (11)

Experimental plots had been fertilized with $0.35 \text{ mol N m}^{-2}$ biweekly throughout the growing season for 3 years, and the grass had grown in height to resemble creekbank grass. Standard errors for three samples are in parentheses.

Previous measurements of DMS emission have shown a high degree of temporal and spatial variability. Using a flow-through system with intact plants, Steudler and Peterson⁹ measured maximum rates of DMS emission during August and September; emission during those months accounted for half their estimate of annual emission. We have not observed seasonal variation in DMSP concentrations in live leaves which might account for this pattern. The increased rate of emission appears to coincide with the decline in leaf biomass late in the growing season and we have measured free acrylic acid ($8 \mu\text{mol g}^{-1}$ dry weight) in moribund and standing dead leaf litter, suggesting that DMS is liberated during leaf senescence, leaving residual acrylic acid.

The spatial variability observed by others^{12,13} may be the result of plants of the same species being in different physiological states, or different plant species dominating neighbouring sites. The low levels of DMSP in other species, even within the genus *Spartina* (Table 1), suggest that DMS emission from areas of marsh dominated by other species will be low. Previous estimates for DMS emission from salt marshes are almost certainly overestimates, as they have been based on fluxes from *S. alterniflora* marshes which were extrapolated to marsh area without considering the plant species. Greater understanding of the mechanisms of DMSP conversion to DMS will enhance our ability to estimate DMS emissions accurately and to establish the significance of salt marshes in global sulphur emissions. Similar metabolism of sulphonium compounds in leaves of

higher plants may represent a major source of terrestrial DMS.

We thank Linda Hare for research assistance, Mike Hardisky for samples and Paul Steudler for comments. This research was supported by NASA.

Received 19 May; accepted 6 October 1987.

1. Andreae, M. O. & Raemdonck, H. *Science* **221**, 744-747 (1983).
2. Lovelock, J. E., Maggs, R. J. & Rasmussen, R. A. *Nature* **237**, 452-453 (1972).
3. Liss, P. S. & Slater, P. G. *Nature* **247**, 181-184 (1974).
4. Maroulis, P. J. & Bandy, A. R. *Science* **196**, 647-648 (1977).
5. Hitchcock, D. *Chemosphere* **3**, 137-138 (1975).
6. Adams, D. F., Farwell, S. O., Robinson, E., Pack, M. R. & Barnesberger, W. L. *Envir. Sci. Technol.* **15**, 1439-1498 (1981).
7. Aneja, V. P., Aneja, A. P. & Adams, D. F. *J. air Pollut. Control Ass.* **32**, 803-807 (1982).
8. Goldberg, A. B., Maroulis, P. J., Wilner, L. A. & Bandy, A. R. *Atmos. Envir.* **15**, 11-18 (1981).
9. Steudler, P. A. & Peterson, B. J. *Nature* **311**, 455-457 (1984).
10. Steudler, P. A. & Peterson, B. J. *Atmos. Envir.* **19**, 1411-1416 (1985).
11. Jorgensen, B. B. & Okholm-Hansen, B. *Atmos. Envir.* **11**, 1737-1749 (1985).
12. Cooper, D. J. *et al. Atmos. Envir.* **21**, 7-12 (1987).
13. de Mello, W. Z. *et al. Atmos. Envir.* **21**, 987-990 (1987).
14. King, G. M., Howes, B. L. & Dacey, J. W. H. *Geochim. cosmochim. Acta* **49**, 1561-1566 (1986).
15. Howes, B. L., Dacey, J. W. H. & Wakeham, S. G. *Limnol. Oceanogr.* **30**, 221-227 (1985).
16. Challenger, F. *Aspects of the Organic Chemistry of Sulfur* (Academic, New York, 1959).
17. Reed, R. H. *Mar. Biol. Lett.* **4**, 173-181 (1983).
18. White, R. H. *J. mar. Res.* **40**, 529-536 (1982).
19. Larher, F., Hamelin, J. & Stewart, G. R. *Phytochemistry* **16**, 2019-2020 (1977).
20. van Diggelen, J., Rozema, J., Dickson, D. M. J. & Broekman, R. *New Phytol.* **103**, 573-586 (1986).
21. Wyn Jones, R. G. & Storey, R. in *The Physiology and Biochemistry of Drought Resistant Plants* (eds Pate, L. G. & Aspinall, D. J.) 171-204 (Academic, Sydney, 1982).
22. Nriagu, J. O., Holdway, D. A. & Coker, R. D. *Science* **237**, 1189-1192 (1987).
23. Andreae, M. O. & Andreae, T. W. *J. geophys. Res.* (in the press).
24. Vairavamurthy, A., Andreae, M. O. & Iverson, R. L. *Limnol. Oceanogr.* **30**, 59-70 (1985).
25. Dacey, J. W. H. & Wakeham, S. G. *Science* **233**, 1314-1316 (1986).
26. Winner, W. E. *et al. Nature* **289**, 672-673 (1981).

Holocene vegetation zonation in the eastern Sahara

J. C. Ritchie* & C. V. Haynes†

* Scarborough College, University of Toronto, Scarborough, Ontario, Canada, M1C 1A4 and Laboratoire de Palynologie, Centre National de la Recherche Scientifique, 34060 Montpellier Cedex, France

† Departments of Anthropology and Geosciences, University of Arizona, Tucson, Arizona 85721, USA

Recent discoveries of fossil-bearing Holocene lake sediments from the eastern Sahara have brought further confirmation and detail to earlier indirect evidence that a major pluvial episode occurred between 9,500 and 4,500 yr BP¹⁻³. The botanical results, mainly palynological, show that savanna and desert grassland occupied regions that today are plantless hyperarid deserts. We present here the first evidence from the entire Saharan region of a latitudinal zonation of Holocene vegetation, extending across a 500-km wide belt in north-west Sudan. We have completed detailed pollen analyses of three of the several sites discovered so far, and the salient features from two, the northernmost and the southernmost, are presented here to demonstrate a steep gradient of past vegetation cover.

The sites consist of moist lacustrine sediment, buried by aeolian sand, in depressions where the present-day watertable is within a few metres of the surface. The sediments are predominantly autochthonous, consisting of alternating units of carbonate precipitates, diatomites and algal gyttjas, with many of the sedimentary units perfectly or almost-perfectly laminated, formed in chemically stratified, meromictic lakes^{1,2}. Consequently, the pollen is of primary origin, and excellently preserved.

The Selima Oasis (21°22' N, 29°19' E), the northernmost site, is in the hyperarid core of the eastern Sahara⁴ (Fig. 1). Holocene sediments were first discovered there by Sandford⁵, and later sampled and preliminarily analysed by Haynes⁶ and Mehringer⁷.

Our pollen samples were taken at 5-cm intervals from a 3.7-m vertical face in a pit (Hole-6 Pit) excavated in beds of moist lacustrine sediment. The radiocarbon chronology (Table 1) is

based on 29 dated levels where charcoal, or algal sapropel, or carbonate were sampled. While predictable discrepancies occur because of the high carbonate content of the former lake, the data fit linear and quadratic regression equations for age/depth with R^2 values of 0.79 and 0.81, respectively. Stratigraphic details of the sediments and radiocarbon dates are described elsewhere².

Pollen-bearing sediment occurred between 1.75 and 3.65 m, and pollen sums of >1,000 were tabulated for each sample. A summary percentage diagram is shown in Fig. 2 with three sections: totals of tropical taxa grouped in their main phytogeographical categories with the total of arboreal taxa; selected common tropical taxa that show continuous representation in the section; and a combined diagram showing arboreal taxa, Chenopodiaceae-Amaranthaceae, Gramineae, Cyperaceae, *Typha* and the sum of all others. Table 2 lists those taxa not shown in Fig. 2, grouped phytogeographically.

The Selima pollen record shows low but continuous frequencies of important indicator taxa, in particular *Acacia*, *Blepharis*, *Commiphora*, *Maerua*, *Salvadora* and *Tribulus*. The tropical taxa are dominated by Saharan and Sahelian types, with low representation of Sudanian elements. The lowest levels (3.50-3.65 m) show high percentages of *Typha*, and this correlated with the presence at the base of the section, and at several of the other sites², of carbonized, matted remains of *Typha*. Haynes² discusses this feature and suggests that the early Holocene rise in groundwater initiated the formation of *Typha* marshes, and later of lakes. From 3.50 to ~2.40 m, Gramineae and Cyperaceae dominate the local taxa, and at 2.40 m there is a sharp increase in Chenopodiaceae-Amaranthaceae. Comparison with modern pollen spectra from elsewhere in the Sahel⁸⁻¹¹, indicate that between ~10,000 and 6,000 yr BP the area immediately surrounding Selima was occupied by a sparsely wooded steppe desert, dominated by *Acacia*, *Commiphora* and *Maerua*, with abundant but fluctuating components of annual perennial herbs and shrubs such as *Tribulus*, members of the Chenopodiaceae-Amaranthaceae complex and *Blepharis*. Modern vegetation analogous to this assemblage occurs throughout the Sahelian zone of the Sudan and Chad^{12,13}. In particular, Vincens⁹ notes that *Acacia-Commiphora* are characteristic pollen types of the shrub-steppe desert in East Africa, although the total arboreal pollen percentage is only 2-4%.

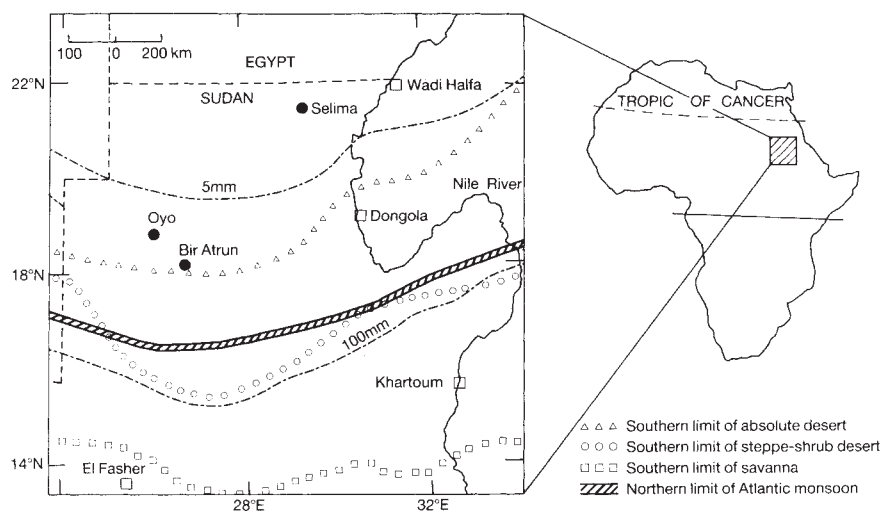


Fig. 1 Map of north-west Sudan showing the locations of the pollen sites (●), the approximate positions of the 5-mm and 100-mm isohyets and the August (maximum northerly) position of surface Atlantic monsoon air flow (after Leroux⁴), approximate vegetation zone boundaries between Saharan desert, Sahelian steppe and savanna and Sudanian woodlands south of 14° N (after White¹⁴).

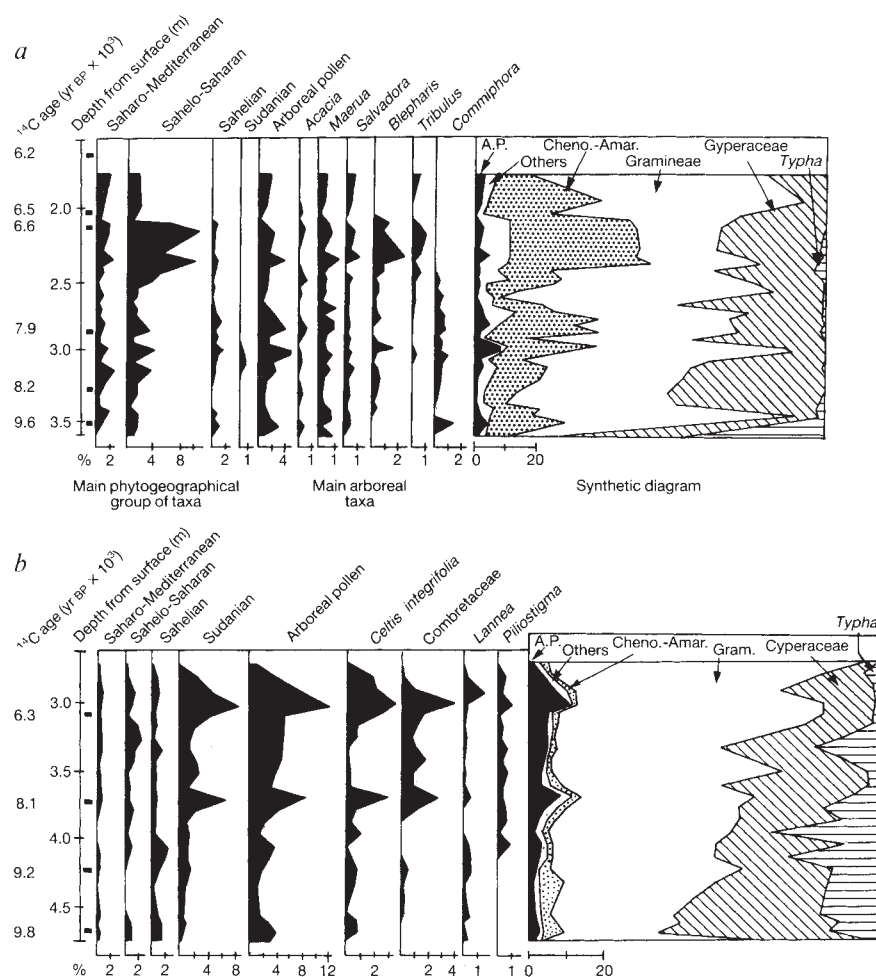


Fig. 2 Summary percentage pollen diagrams from Selima (a) and El'Atrun (b) showing, from left to right on each, radiocarbon ages, depth from surface; frequencies of the four main phytogeographical pollen groups, the total arboreal pollen, the most abundant tropical pollen taxa, and a synthetic diagram showing the four important local taxa (Gramineae, Chen-Amaranthaceae, Cyperaceae and Typha), the arboreal taxa, and the sum of all others.

El'Atrun (18°10' N, 26°39' E) is an oasis 450 km south-west of Selima, near the southern limit of the hyperarid, absolute desert zone¹⁴ (Fig. 1). The sediments occur in a large depression (4 × 8 km) surrounded by bedrock uplands. The surface of the depression consists of aeolian sand containing crystalline aggregates of trona, a commercial evaporite extensively mined by hand. Pollen-bearing sediment occurs between 2.5 m and 4.7 m, ending where an impenetrable bedrock layer was encountered. The sediment was recovered by a modified Livingstone piston corer, with plastic core barrels that were sealed immediately after coring. The sediments consist of finely laminated organic

and carbonate muds. Five samples were radiocarbon dated (Table 1) in addition to one basal core sample taken from a test coring ~30 m from the main pollen corehole. Samples for pollen analysis were taken at intervals between 5 and 10 cm and sums between 1,500 and several thousand were enumerated for each. The results (Fig. 2, in the same format as for Selima) show significant qualitative and quantitative differences from the Selima diagram. At El'Atrun, Saharo-Mediterranean taxa are scarce, while among the tropical elements Sudanian taxa dominate. In particular, the continuous curves of *Celtis integrifolia*, *Combretaceae*, *Lannea* and *Piliostigma* are major differences

Table 1 Radiocarbon dates from the El'Atrun and Selima sediments

Depth from surface (m)	¹⁴ C age (yr BP)	Standard deviation	Material analysed	Radiocarbon laboratory number
El'Atrun				
3.06	6,280	±60	Sapropel	TO-599
3.74	8,060	±70	Sapropel	TO-598
4.21	9,220	±70	Sapropel	TO-600
4.70	7,890	±70	Sapropel	TO-597
4.11*	9,880	±100	Sapropel	TO-181
Selima				
0.41	5,120	±120	Carbonate	A-2267
1.18	4,090	±500	Charcoal	SMU-1024
1.18	6,320	±70	Carbonate	SMU-1022
1.9	7,470	±380	Sapropel	A-2786
2.04	6,460	±90	Carbonate	WSU-2599
2.04	6,510	±120	Sapropel	A-2871
2.11	6,070	±290	Sapropel	AA-102
2.11	6,810	±130	Carbonate	A-2265
2.15	6,620	±170	Sapropel	A-2868
2.15	7,180	±90	Carbonate	WSU-2614
2.91	7,980	±160	Carbonate	A-2257
3.03	7,680	±130	Carbonate	WSU-2610
3.03	8,350	±140	Sapropel	A-2866
3.41	7,730	±210	Charcoal	AA-104
3.41	8,200	±190	Sapropel	A-2874
3.45	7,860	±310	Charcoal	AA-252
3.45	9,850	±90	Carbonate	WSU-2603
3.45	10,280	±220	Sapropel	A-2837
3.53	9,430	±80	Charcoal	TO-415
3.58	9,090	±110	Carbonate	WSU-2601
3.58	9,650	±200	Sapropel	A-2872
3.61	9,340	±200	Sapropel	A-2253
3.61	10,060	±300	Carbonate	A-2252
3.61	8,820	±90	Charcoal	TO-414
3.63	8,880	±70	Charcoal	TO-413
3.63	9,060	±810	Carbonate	A-2250
3.63	9,640	±700	Charcoal	SMU-1223
3.63	9,700	±200	Charcoal	A-2251

El'Atrun samples from cores; Selima samples were taken from a pit. Radiocarbon laboratories: Arizona National Science Foundation Regional Facility for Accelerator Mass Spectrometry (AA), University of Arizona Radiocarbon Dating Laboratory (A), Southern Methodist University, Texas (SMU), Washington State University (WSU) and Isotracer Laboratories, University of Toronto (TO).

* The basal organic material from a core 10 m from the main section.

from the Selima record. Among the local taxa, Chenopodiaceae-Amaranthaceae have consistently low frequencies while Gramineae, Cyperaceae and *Typha* are the dominants. A comparison of the percentage curves of local taxa provides confirmation of significant regional differences between the two sites—at Selima, high values of Chenopodiaceae-Amaranthaceae and low percentages of *Typha*, and the converse at El'Atrun. The role of long-distance pollen transport was insignificant. On the basis of comparison with modern spectra from other Sahelian-Sudanese zones of North Africa⁸⁻¹¹ and descriptions of regional vegetation at nearby uplands (the Ennedi Massif¹², 400 km west-south-west of El'Atrun, and the Darfur uplands¹³, 400 km south-south-west), we suggest that between ~10,000 and 6,000 yr BP, the surroundings of El'Atrun were covered by wooded savanna vegetation very similar to modern communities found in the Ennedi, in Darfur, and throughout the Sahelo-Sudanese vegetation zone of Sudan and adjacent regions¹⁴. These results are very similar to those from the Oyo Depression, 100 km north-west of El'Atrun. Pollen analysis in Saharo-Sahelian zones of North Africa has particular difficulties, of which the most significant is that almost all the tropical taxa dominating the savanna and steppe savannas are low pollen producers, often with large grains that disperse over short distances. As a result, pollen spectra are dominated by such

Table 2 Pollen taxa recorded from both of, or either of Selima and El'Atrun

Saharo-Mediterranean	Sahelian	Sudanese
<i>Artemisia</i>	<i>Balanites</i>	<i>Ampelocissus</i> (A)
Cruciferae	<i>Boerhavia</i>	<i>Cadaba</i> (A)
<i>Ephedra</i>	<i>Borreria</i> (A)	<i>Chrozophora</i>
<i>Erica arborea</i> (S)	<i>Boscia</i>	<i>Hyphaene</i> (A)
<i>Rumex cf. vesicarius</i>	<i>Capparis</i>	<i>Justicia</i>
(S)	<i>Cassia</i> (A)	<i>Lepidagathis</i> (A)
Umbelliferae	<i>Caylusea</i>	<i>Mitragyna</i> (A)
	<i>Chascanum</i>	<i>Securinega</i> (A)
	Convolvulaceae	<i>Sterculia</i> (A)
Saharo-Sahelian	<i>Indigofera</i>	<i>Vernonia</i> (S)
	<i>Lauremburgia</i> (A)	
<i>Abutilon</i>	<i>Merremia</i>	Others
<i>Cleome</i>	<i>Oldenlandia</i>	<i>Ambrosia maritima</i>
<i>Commicarpus</i>	<i>Phyllanthus</i> (A)	Compositae-Asteraceae
<i>Corchorus</i>	<i>Pavonia</i> (A)	Compositae-Fenestratae
<i>Tamarix</i>	<i>Grewia</i>	
<i>Ziziphus</i>		<i>Mitracarpus</i> (A)
		<i>Olea</i>
		Resedaceae
		<i>Trichodesma</i> (S)

Taxa are grouped according to their geographical ranges, based on the analysis of the nearby flora of the Ennedi Massif¹². S, Selima only; A, El'Atrun only. These taxa are in addition to those shown in Fig. 2.

local taxa as Gramineae, Cyperaceae and Chenopodiaceae-Amaranthaceae⁸⁻¹¹.

We suggest that the pollen records from Selima and El'Atrun demonstrate a steep vegetational gradient in the eastern Sahara during an early to mid-Holocene pluvial, with a zonation of wooded savannas at the latitude of El'Atrun and sparsely wooded desert steppes at the latitude of Selima. The northern limit of the sparsely wooded desert steppes was 400 km north of its present-day position, at ~16° N in north-west Sudan, while the northern limit of the Sudanese savanna woodlands, currently at ~14° N in Darfur and Kordofan (Fig. 1), would have been displaced north to at least the latitude of El'Atrun and Oyo (19° N). Obviously more sites will have to be examined to elaborate and bring precision to this reconstruction, but the indication is that the well-defined latitudinal zonation that characterizes the modern vegetation existed also in the early Holocene, displaced northward by at least 4° (or 450 km). These and other analyses from the Saharo-Sahelian belt^{15,16} imply that the steep gradient of summer precipitation (100–440 mm per yr) which occurs from 12 to 17° N latitude was displaced 4 to 5° northward from 10,000 to 5,000 yr BP.

This research has been supported by the National Science and Engineering Research Council of Canada, the NSF, the National Geographic Society, the Geological Survey of Egypt and the Geological and Mineral Resources Department of Sudan. J.C.R. acknowledges the assistance of colleagues at both the Scarborough and Montpellier laboratories.

Received 20 August; accepted 7 October 1987.

- Ritchie, J. C., Eyles, C. E. & Haynes, C. V. *Nature* **314**, 352–355 (1985).
- Haynes, C. V. in *Prehistory of Arid North Africa: Essays in honour of Fred Wendorf* (ed. Close, A. E.) 69–84 (Southern Methodist Univ., Dallas, 1987).
- Pachur, H.-J. & Kröppel, S. *Science* **237**, 298–300 (1987).
- Leroux, M. *Le climat de l'Afrique tropicale* (Champion, Paris, 1983).
- Sandford, K. S. *Quat. J. geol. Soc.* **91**, 323–381 (1936).
- Haynes, C. V. *Geol. Soc. Am. Abstr.* **14**, 511 (1982).
- Mehring, P. J. Jr *Geol. Soc. Am. Abstr.* **14**, 512 (1982).
- Lézin, A. M. thesis, Univ. Marseille (1987).
- Vincens, A. *Rev. Paléobiol.* Vol. spéc. 235–242 (1984).
- Maley, J. *Travaux et Documents de l'ORSTOM* 129 Paris (1981).
- Schultz, E. *Palaeogeography Afr.* **16**, 245–253 (1984).
- Gillet, H. *Le peuplement végétal du massif de l'Ennedi (Tchad)*. B7, 1–206 (1968).
- Quézel, P. *Flora et végétation des plateaux du Darfur Nord-occidental et du Jebel Gourquell (Soudan)* (CNRS-RCP 45, 1969).
- White, F. *The Vegetation of Africa* (UNESCO, Paris, 1983).
- Haynes, C. V., Ritchie, J. C. & Eyles, C. H. *Geochronology of Quaternary Groundwater-Supported Lakes, Northwestern Sudan*, Abstr. (INQUA, 1987).
- Haynes, C. V. & Mead, A. R. *Quat. Res.* **8**, 86–99 (1987).

Roaring by red deer stags advances the date of oestrus in hinds

Karen McComb

Large Animal Research Group, Department of Zoology, Cambridge CB3 0DT, UK

Some male mammals call loudly and repeatedly during the breeding season¹⁻⁵. Although the song of male birds is known to have effects on male-male competition⁶, mate selection and ovulation⁸⁻¹⁴, until now the loud calls of male mammals have been shown to affect only competition between males^{1,15}. Although it has been suggested that loud calling could also serve to attract females^{16,17}, the possibility that it has a direct effect on reproduction in females has not previously been investigated for any mammal. Here I report that roaring in red deer (*Cervus elaphus*) advances ovulation and that harem-holding males can improve their mating success by regular calling.

During the autumn rut, red deer stags gather and defend harems of females, roaring repeatedly as they do so. Roars are low pitched vocalizations (fundamental frequency = 121 ± 33 Hz; mean $\pm$ s.d.) delivered in bouts of one to nine roars, each on a single forced exhalation. Harem-holding stags commonly maintain roaring rates of two roars per min throughout the 24 h, although rates of up to eight roars per min are sustained during roaring contests that precede fights¹. But males also roar when competitors are absent, especially after chasing or herding hinds, or when new hinds join their harem. In several bird species, singing by males has been shown to advance ovulation in females⁸⁻¹⁴. If loud calling in mammals had similar effects, harem-holding stags could increase their mating success by roaring. The most successful stags begin to hold their harems before the onset of the conception peak and harem tenure rarely exceeds 3-4 weeks. By this stage around 20% of hinds have not conceived¹⁸. Consequently, by advancing the date of oestrus, stags would be likely to increase their mating success.

The effect of roaring on ovulation cannot be assessed in free-ranging populations and so was tested experimentally in farmed red deer in New Zealand. Their breeding season normally commences at the beginning of April, six months later than in Britain. Oestrus lasts from 12 to 24 h, and if a female fails to conceive, oestrus will recur at 18 day intervals¹⁹. Calving occurs on average 233 days after conception²⁰.

To test the effect of roaring presented as an isolated stimulus and of roaring combined with other male stimuli (for instance pheromones and chasing), on hind conception dates, three experimental groups were established on 18 March. For 14 days before 1 April, when they were placed with fertile sire stags, these three groups of hinds received different treatments (Fig. 1). One group was exposed to recorded roaring played back through loudspeakers at rates chosen to simulate those normally exhibited by harem-holding stags in a free-ranging population (Fig. 2). To minimize habituation of hinds to playback, roars were selected from a sample of 106 bouts recorded from two different mature stags. The tapes from the two stags were played alternately from day to day. By broadcasting from two speakers separately the direction from which roars were issued was changed every one to six bouts. To determine the effects of roaring combined with other aspects of male behaviour (which include regular disturbance and herding of hinds by the holding stag), a vasectomized male was put with a second group. Vasectomy results in infertility without altering normal rutting behaviour—mean roaring and chasing/herding rates were 43.9 ± 35.6 roars and 2.70 ± 3.70 chases/instances of herding per h, similar to those observed in wild red deer on Rhum¹. A control group was isolated from males and from roaring for the entire treatment period. To identify the effect of male stimuli apart

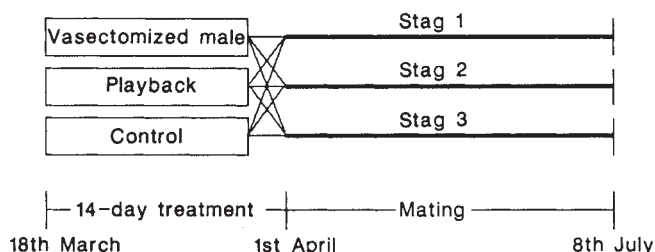


Fig. 1 Experimental design. The 14-day treatments commenced on 18 March, approximately 2 weeks before the onset of the normal breeding season for these hinds. Subsequently, animals were allocated at random to three mating groups. This served to randomize sire effects and ensure that any tendency for hinds to synchronize oestrus within mating groups should obscure the effects of different treatments.

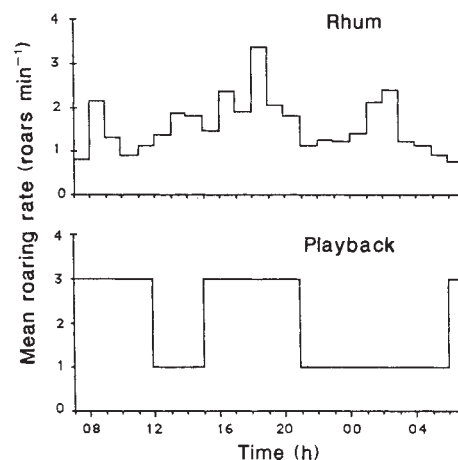


Fig. 2 Mean roars per minute calculated across hour of the day for 10 harem-holding stags in a free-ranging population of red deer on the island of Rhum, Inner Hebrides, Scotland¹, compared to those broadcast to hinds in the playback group. The playback system consisted of a Sony WM-28 stereo cassette player, Tresham Audio SR202 160 W power amplifier and two Tannoy Lynx studio monitor speakers.

from roaring, it would have been ideal to expose a fourth group of hinds to a surgically silenced, vasectomized male, but this was avoided for humanitarian reasons.

At the end of 14 days, the three groups of hinds were each divided into three equal subgroups. These subgroups were then evenly distributed across three fertile sire stags, thus randomizing sire effects across treatment groups. The redistribution also served to counteract the potential problem of oestrus synchrony. Hinds that associate together during the rut may show greater synchrony of oestrus than non-associates²¹. If mating groups were constituted directly from treatment groups, greater within-group synchrony could contribute to between-group differences and erroneously appear as a treatment effect. In this experiment, however, any tendency for animals to synchronize within mating groups should tend to obscure the effects of different treatments.

Conception dates were assessed from calving dates collected the following spring. Sample sizes of hinds completing the experiments were 48 for playback, 36 for vasectomized male and 44 for control groups. For the purposes of analysis, skewness in the distribution of calving dates was removed by log transformation of the data.

Hinds exposed either to roaring or to the vasectomized stag calved earlier than control hinds (Fig. 3). Analysis of variance revealed that of the variables (treatment, sire stag, sex of calf and weight of dam) both treatment and sire stag made significant contributions to the variance in calving date ($F_{2,94} = 6.6159$,

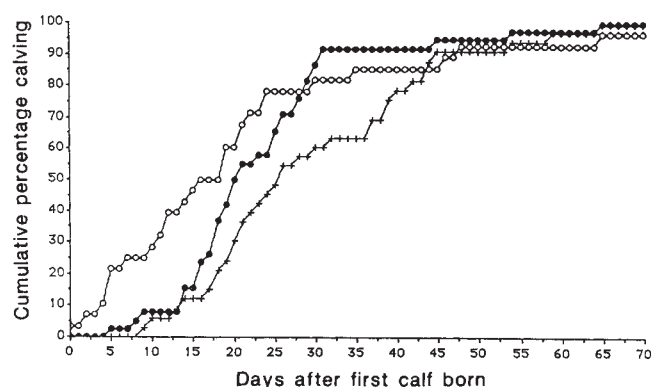


Fig. 3 Cumulative percentage calving in the three groups: control (+), playback (●) and vasectomized (○) male, using values adjusted for variance due to sire stag, sex of calf and sex × sire interaction effects. First calf born on 16 November.

$P < 0.005$ and $F_{2,94} = 9.1802$, $P < 0.005$ respectively) and there was a significant calf sex × sire interaction ($F_{2,91} = 5.2955$, $P < 0.01$). The effect of treatment remained significant when the calving dates were adjusted to take account of the other significant variance components ($F_{2,91} = 5.2884$, $P < 0.01$). When playback and vasectomized-male treatments were compared separately with the control, each had a significant effect. ($F_{1,64} = 5.4289$, $P < 0.025$ and $F_{1,54} = 7.5076$, $P < 0.01$ respectively).

These results show that both the presence of a male and roaring presented as an isolated stimulus can advance conception dates in red deer hinds. The greater effect produced by the presence of a vasectomized male is not surprising, for where male effects on ovulation have been documented in other ungulates^{22–24}, pheromones and male behaviour play an important part²².

If roaring advances ovulation date, selection should favour stags that roar while harem holding, for in doing so they will increase the probability that they mate a hind before losing her to another male. Harem tenure is limited by the rapid decline in body condition of rutting stags. Even the most successful stags, which hold harems during the peak of the rut, lose potential conceptions to inferior males who subsequently hold their hinds^{25,26}. Assuming that failure to roar would retard mean conception date by six days in the free-ranging population on Rhum (the difference in mean conception dates between playback and control groups), the mating success²⁶ of the ten most successful²⁵ stags in a single year would decline by $26.0 \pm 20.9\%$ ($T = 1$, $P < 0.01$, Wilcoxon matched-pairs signed-ranks test). This comparison must be interpreted cautiously as it is not known how frequently stags would have to roar to elicit the response from hinds, but it suggests that the reproductive benefits of roaring could be substantial.

Why should females have been selected to respond to male roars? One explanation is that this could allow them to minimize the possibility that they enter oestrus when no rutting male is in the vicinity. Oestrus commonly lasts for less than 24 h and failure to conceive in one cycle would delay conception for at least 18 days. This would represent a considerable loss in fitness, for calf mortality increases by around 1% for every day the calf is born after the median calving date²⁷. Also, a female's fecundity in the following year declines by 1% for every day in the delay of the conception date of the previous calf²⁸.

This experiment provides the first evidence that male vocalizations can affect the timing of ovulation in female mammals. It raises the possibility that loud calling could have similar functions in other mammal species where calling is not confined to situations in which contests between competing males are imminent^{2–5}.

I thank Tim Clutton-Brock for invaluable help and discussion, Bruce Hegan for supplying facilities at his deer farm, John Milne

for collecting calving dates, Peter Fennessy and Noel Beatson for advice and support, Steve Albon for statistical assistance, Fiona Guinness for access to field data, and colleagues at Cambridge and Edinburgh for improving the manuscript. This work was supported by the British Federation of University Women, the Department of Education for Northern Ireland and Bruce Hegan. Playback equipment was provided by Tannoy of Great Britain.

Received 8 July; accepted 17 November 1987.

1. Clutton-Brock, T. H. & Albon, S. D. *Behaviour* **69**, 145–169 (1979).
2. Shipley, C., Hines, M. & Buchwald, J. S. *Anim. Behav.* **29**, 12–19 (1981).
3. Marler, P. *Behaviour* **152**, 175–197 (1972).
4. Schaller, G. B. *The Serengeti Lion* (Chicago University Press, Chicago, 1972).
5. Harrington, F. H. & Mech, L. D. *Can. J. Zool.* **56**, 2024–2028 (1978).
6. Krebs, J., Ashcroft, R. & Webber, M. *Nature* **271**, 539–542 (1978).
7. Catchpole, C. K. *Trends Ecol. Evol.* **2**, 94–97 (1987).
8. Ficken, R. W., Van Tienhoven, A., Ficken, M. S. & Sibley, F. C. *Anim. Behav.* **8**, 104–106 (1960).
9. Warren, R. P. & Hinde, R. A. *Anim. Behav.* **9**, 64–67 (1961).
10. Brockway, B. F. *Anim. Behav.* **13**, 575–578 (1965).
11. Brockway, B. F. in *Bird Vocalisations* (ed. Hinde, R. A.) 131–158 (Cambridge University Press, 1969).
12. Lehrman, D. S. & Friedman, M. *Anim. Behav.* **17**, 494–497 (1969).
13. Nottebohm, F. & Nottebohm, M. E. *Anim. Behav.* **19**, 313–327 (1971).
14. Kroodsma, D. E. *Science* **192**, 574–575 (1976).
15. Sekulic, R. *Behaviour* **81**, 38–54 (1982).
16. Tyack, P. *Behav. Ecol. Sociobiol.* **8**, 105–116 (1981).
17. Chivers, D. J., Raemaekers, J. J. & Aldrich-Blake, F. P. G. *Folia Primat.* **23**, 1–49 (1975).
18. Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer: Behaviour and Ecology of Two Sexes* (Chicago University Press, 1982).
19. Kelly, R. W., McNatty, K. P. & Moore, G. H. in *Biology of Deer Production* (eds Fennessy, P. F. & Drew, K. R.) 181–184 (The Royal Society of New Zealand, Wellington, 1985).
20. Kelly, R. W. & Moore, G. H. *N.Z. Agric. Sci.* **11**, 179–181 (1977).
21. Iason, G. R. & Guinness, F. E. *Anim. Behav.* **33**, 1169–1174 (1985).
22. Knight, T. W. *Proc. N.Z. Soc. Anim. Prod.* **43**, 7–11 (1983).
23. Chemineau, P. *J. Reprod. Fert.* **67**, 65–72 (1983).
24. Brooks, P. H. & Cole, D. J. A. *J. Reprod. Fert.* **23**, 435–440 (1970).
25. Clutton-Brock, T. H., Albon, S. D., Gibson, R. M. & Guinness, F. E. *Anim. Behav.* **27**, 211–225 (1979).
26. Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. *Reproductive Success: Studies of Individual Variation in Contrasting Breeding Systems* (ed. Clutton-Brock, T. H.) (Chicago University Press, in the press).
27. Clutton-Brock, T. H., Major, M., Albon, S. D. & Guinness, F. E. *J. Anim. Ecol.* **56**, 53–67 (1987).
28. Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *J. Anim. Ecol.* **52**, 367–383 (1983).

Long-term potentiation and NMDA receptors in rat visual cortex

Alain Artola & Wolf Singer

Max Planck Institute for Brain Research, Department of Neurophysiology, Deutscherordenstrasse 46, PO Box 71 06 62, 6000 Frankfurt 71, FRG

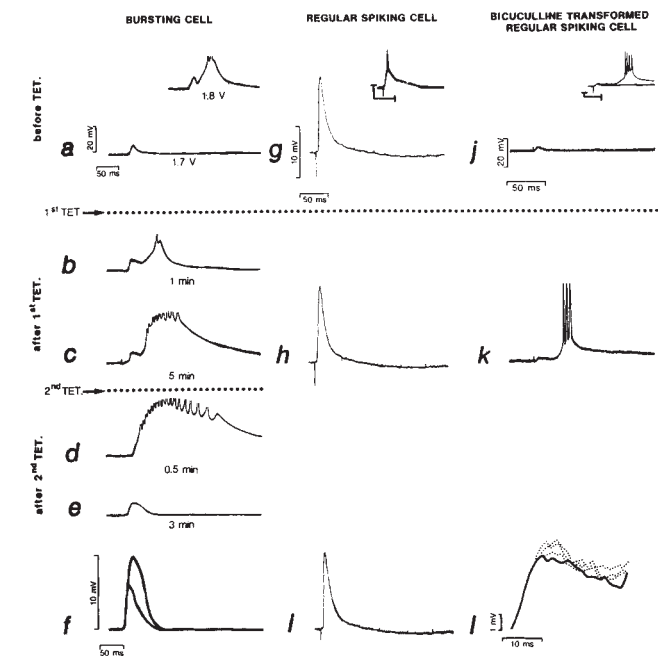
In the hippocampus, which is phylogenetically older than the cerebral neocortex, high frequency stimulation of afferent pathways leads to long-term potentiation (LTP) of synaptic transmission^{1–5}. This use-dependent malleability is of considerable interest because it may serve as a substrate for memory processes. However, in the neocortex, whose involvement in learning is undisputed, attempts to demonstrate LTP have remained inconclusive^{6–8}. Here we use intracellular recording techniques to show that LTP can be induced by high frequency stimulation of the optic radiation in slices of the visual cortex of adult rats. We identify as a necessary prerequisite for the induction of LTP the activation of the membrane channel that is associated with the NMDA (N-methyl-D-aspartate) receptor. Selective blockade of this receptor system with DL-2-amino-5-phosphonovalerate consistently prevents LTP as in most hippocampal pathways^{9–11}. In most cortical neurons the activation of the NMDA mechanism and hence the induction of LTP in these experiments requires a concomitant reduction of GABAergic inhibition by low doses of the GABA_A antagonist bicuculline. This indicates that in the neocortex the activation threshold of the NMDA-mechanism and consequently the susceptibility to LTP, are strongly influenced by inhibitory processes.

Fig. 1 Effect of tetanic stimulation on orthodromic responses of a bursting cell (*a-f*), of a regular-spiking cell (*g-i*) and of bicuculline transformed regular-spiking cells (*j-l*). *a*, Control response of the bursting cell before tetanus. Low intensity white-matter stimulation evokes an e.p.s.p. (peak latency range from 5–10 ms) whose amplitude increases gradually with stimulus strength. At higher stimulus intensities a delayed depolarization appears (*a* inset) that gives rise to a high frequency burst of spikes (2–5). This delayed depolarization has all-or-none threshold (here 1.8 V) and amplitude characteristics. Further increasing stimulus intensity only reduces burst latency. *b* and *c*, Responses to test stimuli (1.7 V, 200 μ s) 1 min and 5 min after the first tetanus showing post-tetanic potentiation (PTP) (*b*), and after a subsequent post-tetanic depression (not shown) the beginning of LTP (*c*). The shape of the PTP response is the same as that of the suprathreshold response (*a* inset) whereas the LTP response shows a dramatic increase in the burst-generating depolarization. *d* and *e*, Responses to test stimuli (1.7 V, 200 μ s) obtained 30 s and 3 min after a second tetanus, delivered 63 min after the first. Note the further amplitude enhancement and latency reduction of the burst-generating depolarization as early as 30 s after the tetanus (*d*). The response in *e* remained subthreshold because test stimuli were applied at 0.03 Hz so that every other stimulus fell into the prolonged refractory period that followed each burst response. *f*, Superimposed responses *a* and *e* at magnified scale. Note the amplitude increase of the test e.p.s.p. after LTP (resting membrane potential $V_{mr} = -73$ mV). *g* and *j* Insets, superimposed orthodromic responses to one sub- and one just suprathreshold white-matter stimulus in a normal (*g*) and a transformed regular-spiking cell (*j*). The orthodromic response in *j* is similar to that of the bursting cell and there was never a spontaneous burst. *g*, *h*, Averages of ten successive responses of a regular spiking cell (2 V, 200 μ s, 0.03 Hz) immediately before (*g*) and 15 min after the tetanus (*h*). *i*, Superimposed traces (*g* and *h*). ($V_{mr} = -65$ mV). *j* and *k*, Responses to test stimuli (1.3 V, 200 μ s) before (*j*) and 11 min after (*k*) tetanic stimulation. Note that the large burst (*k*) does not differ from the normal suprathreshold response (*j* inset) ($V_{mr} = -67$ mV). *l*, Averages of 10 e.p.s.ps measured before (continuous line) and after tetanus (dotted lines) at times indicated by triangles in trace *a* of Fig. 3. E.p.s.p. amplitude is consistently enhanced after tetanus ($V_{mr} = -76$ mV). All tetani consisted of five, 2 s long stimulus trains (50 Hz) delivered at 10 s intervals. Stimulus intensity during the tetani was the same as that of the test stimulus for the bursting and the regular-spiking cell and twice that of the test stimulus for the transformed regular-spiking cell. Baseline excitability was assessed with sub-threshold test stimuli applied for 10–20 min before the tetanus at a frequency of 0.03 Hz. Spikes in *a-d*, *g* inset and *k* are truncated and depolarizations underlying the bursts in *c*, *d*, *j* inset and *k* are cut off before V_{mr} is reached. Calibrations are the same for *a-e*, for *g-i* and for *j-k*. Calibrations for *g* and *j* insets are respectively 10 mV and 50 ms and 20 mV and 50 ms.

Methods. Adult albino rats (120–250 g) were decapitated and after quick removal, brains transferred to cold (5 °C), oxygenated (95 % O₂, 5 % CO₂) incubation medium (NaCl 124 mM, KCl 5 mM, NaHPO₄ 1.25 mM, MgSO₄ 2 mM, CaCl₂ 2 mM, NaHCO₃ 26 mM, D-glucose 10 mM, and H₂O₂ 0.002 %). Subsequently, 350 μ m thick vibratome slices of the visual cortex were cut at 20° to the frontal plane, to be as parallel as possible to the trajectories of the radiation fibers. After incubation for 1 h at room temperature, for equilibration, slices were transferred to the recording chamber (submerged type) and superfused at a rate of 4.5 ml min⁻¹ with warm medium (30 °C). Intracellular recordings were made with 3 M (mdar) potassium acetate-filled micropipettes (resistance 80–120 M Ω). Orthodromic responses were elicited by stimulating the underlying white matter with a bipolar tungsten electrode. In some experiments, e.p.s.p. amplitudes were quantified by averaging 10 successive responses. To assess AP5 effects on the orthodromic response of regular-spiking cells, measurements were made three times at 10 min intervals (30 responses) before AP5 application, 5, 10, 20, and 30 min after the onset of the AP5 perfusion (40 responses) and 10, 20 and 30 min after washout (30 responses). In six out of eleven cases, control PSPs (before AP5 application) showed linear decrease in amplitude (mean $\pm$ S.D. = 2.95 ± 1.35 % per 10 min) that continued during AP5 perfusion and washout. To assess accurately AP5 effects, the expected amplitudes of control responses at time of perfusion were extrapolated assuming a linear regression.

Intracellular recordings were obtained from 85 neurons in layers II to IV of rat visual cortex slices. Most of the cells (92%, $n = 78$) had properties similar to those of the recently described 'regular-spiking' cells in the guinea pig cortex and only a small fraction (8%, $n = 7$) resembled 'bursting' cells (for description see Fig. 1 and refs 12 and 13). The two cell types possessed similar membrane properties (Student *t*-test). In regular-spiking cells the resting membrane potential was -72.2 ± 4.0 mV (mean $\pm$ s.d.) and the input membrane resistance was 42.8 ± 18.1 M Ω , whereas in bursting cells the resting potential was -72.5 ± 4.5 mV ($n = 6$) and the membrane resistance 37.0 ± 5.7 M Ω ($n = 4$).

High frequency stimulation had been applied to the white matter and induced LTP in all the bursting cells tested ($n = 5$). Stimulus intensity was adjusted so that a single stimulus elicited a sub-threshold excitatory post-synaptic potential (e.p.s.p.) and the LTP-inducing train produced a brief burst discharge and a sustained depolarization due to temporal summation. As in the hippocampus⁴, the tetanus frequency was critical. Stimulation at 10 Hz was ineffective, at 20 Hz it induced LTP in one cell and at 30–50 Hz it caused LTP in the remaining four cells. Typically, as in the hippocampus^{3,4}, post-tetanic excitability changes were biphasic. Just after the tetanus the amplitude of the test e.p.s.p. was reduced and after 4.5–5.5 min it consistently triggered burst discharges. These bursts were always larger than either those obtained with supra-threshold single shocks before potentiation (compare Fig. 1*a* inset and *c*) or those observed in one cell during transient (1.5 min) post-tetanic potentiation (compare Fig. 1*a* inset and *b*). This suggests that the enhancement of the burst is a specific feature of LTP in this cell type. In all five bursting cells this response enhancement persisted



throughout the observation period which ranged from 30–90 min. If tetanic stimulation was repeated once LTP had been established, post-tetanic depression was shorter and the subsequent response enhancement was increased (Fig. 1*d*). When the time between two test stimuli was reduced below a critical interval ranging from 8–25 s for different cells, the second stimulus failed to trigger a burst discharge (Fig. 1*e*). This made it possible to measure the e.p.s.p. amplitude in the cell shown in Fig. 1 in which the e.p.s.p. amplitude was particularly high. LTP was associated with an enhancement of the e.p.s.p. amplitude, particularly of its late phase (Fig. 1*f*). This enhancement probably accounts for the emergence of burst discharges in response to previously subthreshold test stimuli.

None of the ten regular-spiking cells tested under the same conditions (see above and Fig. 1 legend) exhibited any sign of LTP. Response amplitude remained within 98.8 ± 7.9 % of the control 10–15 min after the tetanus (Fig. 1*g-i*). But we discovered that regular spiking cells could assume the response patterns of bursting cells when exposed to low doses of the GABA_A antagonist bicuculline and then became susceptible to LTP. Bath application of very low doses of bicuculline (0.25–0.5 μ M) had no measurable effect on steady state membrane properties but, after treatment for one hour, the orthodromic responses of regular-spiking cells were indistinguishable from those of bursting cells. High frequency stimulation of the white matter consistently induced LTP in these transformed regular-spiking cells. In seven out of eight cells, test stimuli that were sub-threshold before the tetanus (Fig. 1*j*) evoked burst discharges after tetanization (Fig. 1*k*). Concurrently the e.p.s.p. amplitude was enhanced (Fig. 1*l*). In all cells these signs of

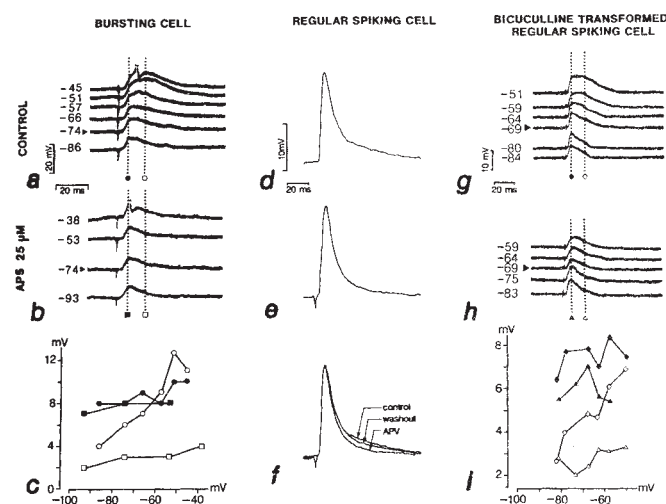


Fig. 2 Responses of a bursting cell (*a-c*), a regular-spiking cell (*d-f*) and a bicuculline ($0.5 \mu\text{M}$) transformed regular-spiking cell (*g-i*) to subthreshold white-matter stimulation before (*a, d, g*) and during (*b, e, h*) bath application of $25 \mu\text{M}$ AP5. *a, b, g* and *h*, Individual responses at different levels of membrane potential V_m . V_{mr} is arrowed. Changes in V_m are induced by long current pulses, synaptic responses being elicited more than 100 ms after pulse onset at rates of 0.1–0.066 Hz. Spikes elicited at depolarized V_m in *a* and *b* are truncated. *d* and *e*, Averages of 10 responses of the regular-spiking cell at V_{mr} , before (*d*) and during AP5 (*e*). *f*, Averages (*d* and *e*) are superimposed with an average obtained 30 min after washout of AP5 ($V_{mr} = -72 \text{ mV}$). Note that in this regular-spiking cell AP5 causes a slight, reversible, increase of the e.p.s.p. decay rate. *c* and *i*, Graphical representations of voltage-dependent amplitude changes of early (filled symbols) and late (open symbols) components of synaptic responses of the bursting cell (*c*) and the bicuculline-transformed regular-spiking cell (*i*) before and during AP5 treatment. Abscissa: V_m , ordinate: PSP amplitude. PSP amplitudes measured at the times indicated by dotted lines in *a, b, g* and *h* (7.5 and 18 ms after stimulation in *c* and 7.5 ms and 20 ms in *i*). These delays correspond to the respective peaks of depolarizations at extreme negative and positive V_m s. In the bursting cell (*c*) and transformed regular-spiking cell (*i*) early components ($\bullet, \blacklozenge$) hardly change whereas late components ($\circ, \diamond$) increase drastically with membrane depolarization. AP5 has only little effect on the early component ($\blacksquare, \blacktriangle$) but reduces the late one ($\square, \triangle$) and abolishes its voltage dependence. The AP5 susceptible e.p.s.p. component (difference between empty symbols in *c* and *i*) increases with membrane depolarization. Calibrations are the same for *a* and *b, d, e* and *f* and *g* and *h*.

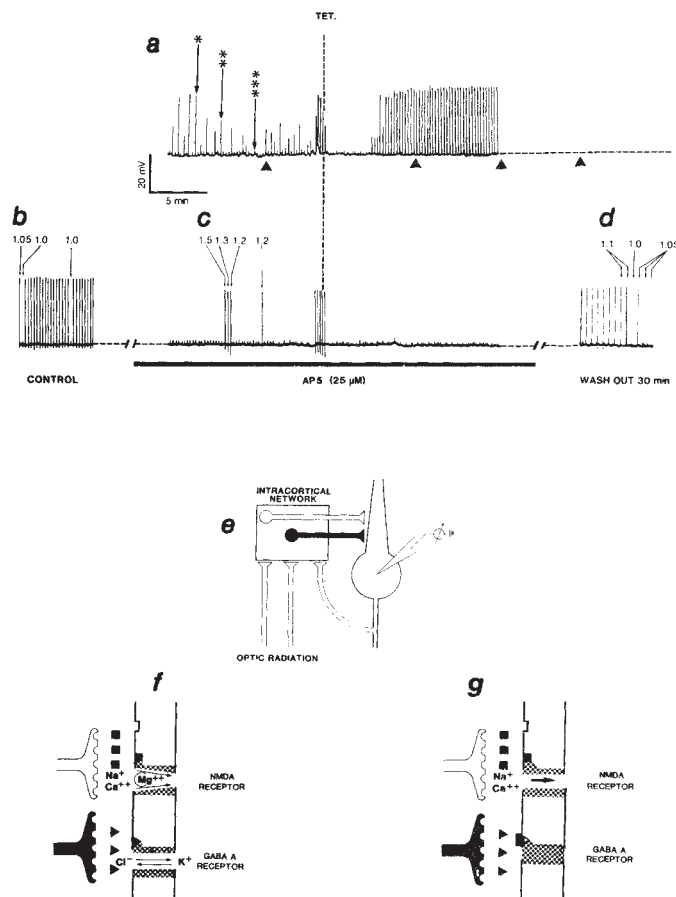


Fig. 3 Effects of tetanic stimulation on bicuculline ($0.5 \mu\text{M}$) transformed regular-spiking cells. *a*, Chart recorder protocol of responses to test stimuli (1.67 V , $200 \mu\text{s}$, 0.06 Hz), 1 h after starting bicuculline perfusion ($0.5 \mu\text{M}$). Before the tetanus, test stimuli elicit an e.p.s.p. (***) which is occasionally followed by delayed depolarization (**) or burst of spikes (*). After tetanic stimulation and a transient post-tetanic depression, test stimuli consistently elicited burst discharges ($V_{mr} = -76 \text{ mV}$). *b-d*, Chart recorder protocol of responses to test stimuli in another cell. *b*, One hour after starting bicuculline perfusion the cell responds with a burst discharge to suprathreshold white-matter stimulation (1.1 V , $200 \mu\text{s}$, 0.06 Hz). *c*, The same cell after bath application of AP5 ($25 \mu\text{M}$). As in the six other cells tested in this way, AP5 raised the threshold (to 1.2 V) and shortened burst duration. Before tetanization, stimulus intensity was set to just subthreshold levels (1.1 V , $200 \mu\text{s}$, 0.06 and 0.03 Hz from 3.5 min before tetanus) for repeated assessment of control response stability. After tetanus no burst appeared. *d*, The stimulus intensities required to elicit suprathreshold responses (see numbers over respective responses in *b* and *d*) show that threshold values tested 30 min after washout of AP5 did not differ from those measured before AP5 application ($V_{mr} = -69 \text{ mV}$). All tetani consisted of five, 2 s stimulus trains (50 Hz) delivered at 10 s intervals. Stimulus intensity during tetani was twice that of test stimuli. Calibrations in *a* apply also to *b, c* and *d*. In *b, c* and *d* stimulus intensity is indicated over the corresponding response when it differs from that in the legend. *e*, Schematic and simplified wiring diagram indicating that white-matter stimulation leads to coactivation of excitatory and inhibitory inputs to the recorded cortical cell. *f* and *g*, We assume that inhibitory and excitatory synapses are in close proximity, the former acting through GABA and GABA $_A$ receptors, the latter through an excitatory amino-acid (EAA), the NMDA receptor and possibly an additional receptor for EAA (?). *f*, With coactivation of excitatory and inhibitory inputs, increased Cl^- and K^+ ion conductance of the GABA channel reduces membrane depolarization and prevents removal of the Mg^{2+} ion block from the NMDA channel. The efficacy of the excitatory synapse is reduced and LTP does not occur. *g*, When the efficacy of the GABAergic synapse is reduced by bicuculline, depolarization caused by the excitatory synapse is sufficient to remove the Mg^{2+} ion block, the NMDA channel becomes permeable to Ca^{2+} ions, efficacy of the excitatory synapse increases and LTP can be induced.

potentiation persisted throughout the recording periods (up to 2 h). There were, however, some differences in the manifestation of LTP in the two cell types. First, in transformed regular-spiking cells the shape of the burst generating depolarization was not markedly modified after LTP (compare Fig 1j inset and k). Second, to be effective in these cells the intensity of the tetanus had to be raised, usually to twice the subthreshold test level. To control for the possibility that raising the intensity of tetanic stimulation, rather than bicuculline treatment, was responsible for the induction of LTP in transformed regular-spiking cells we tested ten untreated regular-spiking cells with high intensity tetani. We obtained no evidence for LTP. The e.p.s.p. amplitude to the subthreshold test stimulus remained unchanged 15–20 min after the tetanus ($99.7 \pm 4.5\%$ of the control).

Untreated regular-spiking cells and natural bursting cells differed in the extent to which the e.p.s.ps elicited by white-matter stimulation were influenced by changes of membrane potential and by bath application of the NMDA antagonist DL-2-amino-5-phosphonovalerate (AP5). In bursting cells ($n = 2$), membrane depolarization enhanced a delayed component (time to peak > 20 ms) of the e.p.s.p. (Fig. 2a) and addition of AP5 (25 μ M) completely suppressed this delayed e.p.s.p. component (Fig. 2b). AP5 had no effect on the steady-state membrane properties and did not affect the early e.p.s.p. component. The amplitude of the late AP5-sensitive e.p.s.p. component changed from 3 mV at resting potential to 10 mV at 20 mV more positive (Fig. 2c). In regular spiking cells ($n = 11$), AP5 also reduced a late e.p.s.p. component (amplitude measured at a delay of 22 ms see below). At resting potential this reduction was only 2.0 ± 0.7 mV (Fig. 2d–f) and, unlike in bursting cells, this late AP5-sensitive component of the e.p.s.p. did not increase with membrane depolarization (data not shown). We attribute this to the concomitant short-latency inhibitory post-synaptic potential (i.p.s.p.) which had a peak latency similar to the late e.p.s.p. component (22.8 ± 4.5 ms, $n = 22$). In transformed regular-spiking cells where this short-latency i.p.s.p. has been blocked by bicuculline, the e.p.s.p. amplitude increased markedly with membrane depolarization. As in bursting cells, this increase was particularly pronounced for a late component (time to peak = 21.4 ± 5.0 ms, $n = 15$) (Fig. 2g). Bath application of AP5 (25 μ M) only slightly reduced (by $14.6 \pm 11.0\%$ at resting potential $n = 5$) the amplitude of the short-latency component of the e.p.s.p. but completely blocked the late component (Fig. 2h). The average amplitude of this late AP5 susceptible e.p.s.p. component was 2.6 ± 1.2 mV at a resting potential of -69.4 ± 3.6 mV, and increased significantly with membrane depolarization ($p < 0.01$) (Fig. 2i). Thus, in regular-spiking cells, bicuculline unmasks a delayed e.p.s.p. which has similar properties to the late e.p.s.p. in natural bursting cells and is mediated by NMDA receptors.

The fact that bicuculline treatment unmasked a NMDA-receptor dependent e.p.s.p. in regular-spiking cells and facilitated the occurrence of LTP suggests an involvement of NMDA-receptor mediated processes in LTP. To test this hypothesis we applied 25 μ M AP5 to four bicuculline transformed regular-spiking cells before tetanic stimulation. Even when we augmented test stimulation intensities to just below threshold and applied high frequency trains (50 Hz) of suprathreshold stimuli (twice the intensity of the test stimulus), we failed to enhance subsequent responses to test stimuli (Fig. 3b–d).

The differences between the orthodromic responses of bursting and regular-spiking cells are mainly due to the reduction of a delayed NMDA-receptor mediated e.p.s.p. component in the latter. The effect of bicuculline treatment suggests that this reduction is due to GABAergic inhibition. We propose that inhibition, either through its hyperpolarizing or its shunting effect, prevents the removal of the voltage-dependent Mg^{2+} block of the ionophore associated with the NMDA receptor^{14,15} (Fig. 3e–g). This interpretation is compatible with the evidence that either reduction of GABAergic inhibition, or removal of Mg^{2+}

ions increase NMDA-receptor dependent e.p.s.p. components in hippocampus^{16–19} and somatosensory cortex²⁰. Assuming that high frequency stimulation attenuates i.p.s.ps¹³, or overrides their effect through temporal summation of e.p.s.ps, our interpretation also explains why such stimulation unmasks AP5-sensitive e.p.s.p. components in the hippocampus²¹.

Our results suggest three major conclusions. First, LTP can be obtained in cortical neurons in a similar way as in the hippocampus^{2–5}, corroborating and extending previous indications for LTP in the cortex^{6–8}. Second, the activation of the NMDA-receptor dependent component of the e.p.s.p. is a necessary condition for the induction of LTP in the cortex. Third, GABAergic inhibition very effectively controls the activation of the NMDA mechanism. This implies that the susceptibility of individual neurons to undergo LTP is crucially dependent on the balance between excitatory and inhibitory inputs, and hence on the architecture of the embedding neuronal network. We propose that the failure of high frequency and high intensity tetani to induce LTP in regular-spiking cells under normal conditions is due to strong inhibitory input. This interpretation is consistent with the finding that blockade of inhibition facilitates LTP in hippocampus²².

The activation of the NMDA mechanism is associated with postsynaptic Ca^{2+} ion fluxes through the NMDA-receptor dependent ionophores^{23,24} and probably also through the voltage-dependent Ca^{2+} ion channels that are probably activated during the tetanus-induced burst discharges²⁵. Thus, as for hippocampal LTP²⁶, cortical LTP could depend on a surge of intracellular free Ca^{2+} ions. The finding that LTP in the cortex requires the activation of the NMDA mechanism and hence has a threshold that is only reached under special activation conditions also suggests interesting parallels with use-dependent long-term modifications of synaptic transmission which occur in the developing visual cortex in response to visual experience. These adaptive processes are also associated with the activation of Ca^{2+} ion fluxes²⁷, have thresholds²⁸ that are related to NMDA mechanisms²⁹ and are reached only if there is sufficient cooperativity between retinal input and additional internally generated gating signals (for review see ref. 30).

We thank Dr C. M. Gray for critically reading the manuscript, R. Ruhl for graphic assistance and C. Ziegler for technical assistance.

Received 22 July; accepted 26 October 1987.

- Bliss, T. V. P. & Lomo, T. *J. Physiol.* **232**, 331–356 (1973).
- Schwartzkroin, P. A. & Wester, K. *Brain Res.* **89**, 107–119 (1975).
- Alger, B. E. & Teyler, T. J. *Brain Res.* **110**, 463–480 (1976).
- Dunwiddie, T. & Lynch, G. J. *Physiol.* **276**, 353–367 (1978).
- Andersen, P., Sundberg, S. H., Sveen, O., Swann, J. W. & Wigström, H. *J. Physiol.* **302**, 463–482 (1980).
- Komatsu, Y., Toyama, K., Maeda, J. & Sakaguchi, H. *Neurosci. Lett.* **26**, 269–274 (1981).
- Lee, K. S. *Brain Res.* **239**, 617–623 (1982).
- Bindman, L. J., Meyer, T. & Pockett, S. J. *J. Physiol.* **386**, 90P (1987).
- Collingridge, G. L., Kehl, S. J. & McLennan, H. J. *Physiol.* **334**, 33–46 (1983).
- Wigström, H. & Gustafsson, B. *Neurosci. Lett.* **44**, 327–332 (1984).
- Harris, E. W., Ganong, A. H. & Cotman, C. W. *Brain Res.* **323**, 132–137 (1984).
- McCormick, D. A., Connors, B. W., Lighthall, J. W. & Prince, D. A. *J. Neurophysiol.* **54**, 782–806 (1985).
- Connors, B. W., Gutnick, M. J. & Prince, D. A. *J. Neurophysiol.* **48**, 1302–1320 (1982).
- Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462–465 (1984).
- Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. *Nature* **309**, 261–263 (1984).
- Herron, C. E., Williamson, R. & Collingridge, G. L. *Neurosci. Lett.* **61**, 255–260 (1985).
- Dingledine, R., Hynes, M. A. & King, G. L. *J. Physiol.* **380**, 175–189 (1986).
- Hablit, J. J. & Langmoen, I. A. *J. Neurosci.* **6**, 102–106 (1986).
- Herron, C. E., Lester, R. A. J., Coan, E. J. & Collingridge, G. L. *Neurosci. Lett.* **60**, 19–23 (1985).
- Thomson, A. M. *J. Physiol.* **370**, 531–549 (1986).
- Herron, C. E., Lester, R. A. J., Coan, E. J. & Collingridge, G. L. *Nature* **322**, 265–268 (1986).
- Wigström, H. & Gustafsson, B. *Nature* **301**, 603–604 (1983).
- MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J. & Barker, J. L. *Nature* **321**, 519–522 (1986).
- Ascher, P. & Nowak, L. *J. Physiol.* **377**, 35P (1986).
- Wong, R. K. S. & Prince, D. A. *Brain Res.* **159**, 385–390 (1978).
- Lynch, G., Larson, J., Kelso, S., Barrionuevo, G. & Schottler, F. *Nature* **305**, 719–721 (1983).
- Geiger, H. & Singer, W. *Expl Brain Res.* **54**, 256–270 (1986).
- Greuel, J. M., Luhmann, H. J. & Singer, W. *Dev Brain Res.* **34**, 141–149 (1987).
- Kleinschmidt, A., Bear, M. F. & Singer, W. *Science* **238**, 355–358 (1987).
- Singer, W. In *The neural and molecular bases of learning* (eds Changeux, J.-P. & Konishi M.) 301–336 (Wiley, Chichester, 1987).

Synergism of inositol trisphosphate and tetrakisphosphate in activating Ca^{2+} -dependent K^+ channels

A. P. Morris, D. V. Gallacher*, R. F. Irvine† & O. H. Petersen

MRC Secretory Control Research Group, The Physiological Laboratory, University of Liverpool, PO Box 147, Liverpool, L69 3BX, UK

† AFRC Institute of Animal Physiology and Genetics Research, Department of Biochemistry, Babraham, Cambridge CB2 4AT, UK

Inositol 1,4,5-trisphosphate (Ins P_3) is a second messenger releasing intracellular Ca^{2+} into the cytosol¹⁻⁵. It has recently been proposed that inositol 1,3,4,5-tetrakisphosphate (Ins P_4), which is formed from Ins P_3 by Ins P_3 -3-kinase (ref. 6), acts with Ins P_3 as a second messenger by promoting extracellular Ca^{2+} entry⁷. It has been suggested that Ins P_3 itself can act to stimulate Ca^{2+} uptake from the extracellular fluid, although a physiological function for Ins P_4 was not excluded^{8,9}. Transmembrane currents can now be measured in single cells by voltage clamping under conditions where the intracellular perfusion fluid can be changed several times during individual experiments^{10,11}. We have used this method to test the effects of Ins P_3 and Ins P_4 on the Ca^{2+} -activated K^+ current^{12,13}, and now show that neither Ins P_3 alone nor Ins P_4 alone can activate a sustained current, whereas Ins P_3 and Ins P_4 in combination evoke a sustained increase in Ca^{2+} -activated K^+ current which is dependent on external Ca^{2+} .

In the lacrimal acinar cells used for this study acetylcholine (ACh) evokes sustained membrane hyperpolarization due to a marked and sustained increase in outward K^+ current, mediated by high-conductance Ca^{2+} - and voltage-activated K^+ channels¹⁴⁻¹⁸. The activation of K^+ current by ACh is only sustained in the presence of extracellular calcium. Figure 1 shows recordings of whole-cell currents from single isolated mouse lacrimal acinar cells. At the holding potential of -40 mV (in some experiments -35 or -50 mV), which is close to the normal resting potential^{14,16}, ACh evokes an outward current response which is substantially enhanced at a membrane potential of zero mV. At more negative potentials (-75 to -90 mV), an inward current response is observed. These currents have been well-characterized: the ACh-evoked outward current is carried by K^+ through high-conductance (200–300 pS) voltage- and Ca^{2+} -activated K^+ channels^{17,18,26}, whereas the inward current is carried by Cl^- through Ca^{2+} -dependent Cl^- channels with a low unit conductance (1–2 pS) (refs 26–28). Figure 1 shows that the effect of continued ACh stimulation is reversible and can be blocked either by introducing a high concentration (10 mM) of EGTA into the cell or by removing external Ca^{2+} . Readmission of extracellular Ca^{2+} during sustained ACh exposure brings back the full response. On the other hand, removal and readmission of external Ca^{2+} in the absence of ACh stimulation has no effect on the magnitude of the transmembrane currents. The effect of ACh can also be reversibly mimicked by introducing a solution of high Ca^{2+} concentration (about $1 \mu\text{M}$) into the cell. The results shown in Fig. 1 are very similar to those from a study of cholecystokinin-evoked Ca^{2+} -dependent K^+ currents in pig pancreatic acinar cells²⁹, but are more direct and convincing as a result of now being able to change the intracellular ion composition during individual experiments.

Figure 2 shows that neither Ins P_3 nor Ins P_4 alone evoked any clear current response in the same lacrimal acinar cell. But when Ins P_4 was added on top of Ins P_3 , a clear and sustained increase in the outward current at 0 mV was obtained. Wash-out of the two inositol phosphates reduced the currents towards the

control level. ACh could thereafter evoke a response of similar magnitude to that induced by combination of the two inositol phosphates. The ACh-evoked outward current was subsequently reversibly abolished by barium ion, a K^+ -channel blocker²¹.

Figure 3 shows results from another cell in which Ins P_3 only evoked a small increase in outward current that declines to pre-stimulus levels within the first minute, even in the continued presence of Ins P_3 . When Ins P_4 is subsequently applied with Ins P_3 still present there is a clear outward current response, as in Fig. 2. The maximal current amplitude is at three minutes, but increased outward currents are still recorded after five minutes. At this point, as seen in Fig. 3, the removal of extracellular calcium results in an abrupt decrease in the amplitude of the outward currents at 0 mV to pre-stimulus levels. This effect of calcium removal is readily reversed, and readmission of extracellular calcium restores the response which persists for the duration of the experiment.

In view of the well-established messenger function for Ins P_3 (refs 1–5), the relatively small response it evoked (Fig. 3), even the complete absence of any effect (Fig. 2), was surprising. But responses in perfused cell systems are dependent on experimental conditions (see legend to Fig. 4) and the protocols used (for example, if Ins P_3 is present in the pipette while the whole-cell current recording mode is established, transient increases in current are obtained⁹ which last longer than those we report using the internal perfusion). Although we found some variation between different preparations of lacrimal acinar cells in responsiveness to Ins P_3 , the marked synergism evident on adding Ins P_3 and Ins P_4 together was highly reproducible. The range of responses to Ins P_3 is illustrated in Fig. 4. The top trace represents the typical small and transient effect induced by $10 \mu\text{M}$ Ins P_3 . The middle trace is exceptional and shows the largest and longest response to $10 \mu\text{M}$ Ins P_3 obtained in our experiments, but is nevertheless transient, the currents returning to pre-stimulus levels within 3 minutes. As Ins P_3 is constantly being delivered to the cell interior from the pipette (which has an enormous volume compared to the single cell) it is unlikely that the transient or absent (Fig. 2) response to Ins P_3 can be explained by its enzymatic breakdown. The Ins P_3 concentration used is maximal or supramaximal for Ca^{2+} release in other permeabilized epithelial cells^{1,3}, but Ins P_3 at $100 \mu\text{M}$ was also tested. The responses to this very high dose of Ins P_3 (shown in the bottom trace of Fig. 4) were not enhanced over the $10 \mu\text{M}$ concentration. It is therefore unlikely that the effect of Ins P_4 addition is due to prevention of Ins P_3 hydrolysis by competition for the phosphatase which hydrolyses both molecules³⁰.

It is interesting that ACh can evoke full current responses in cells where Ins P_3 alone has little or no effect (Fig. 2). Because the ACh response is blocked by internal Ca^{2+} chelation and can be mimicked by internal Ca^{2+} application, the simplest interpretation is that the effect is primarily caused by an increase in internal Ca^{2+} concentration evoked by Ins P_3 -mediated Ca^{2+} release. The sustained ACh response and its dependence on external Ca^{2+} could be explained if Ins P_4 were formed from Ins P_3 . But if Ins P_4 can be formed from Ins P_3 during ACh stimulation, then why not in the absence of stimulation? The activity of the specific Ins P_3 3-kinase (ref. 6) could be regulated by muscarinic receptor activation. Ca^{2+} is a potent stimulator of soluble Ins P_3 3-kinase (refs 31, 32), but this cannot entirely explain our results as little soluble activity would remain in these perfused cells. But other control mechanisms of a membrane-bound activity could be operative such as regulation by a receptor-coupled G protein or by protein kinase C.

Our results on the control of the outward Ca^{2+} -dependent K^+ currents are clear, but the control of the inward Cl^- currents is more complex. The inward Cl^- current responses to ACh stimulation are larger than those evoked by Ca^{2+} or by $\text{Ins P}_3/\text{Ins P}_4$, and perhaps involve an additional regulatory pathway such as protein kinase C stimulation. To explain acinar fluid secretion it is necessary to have simultaneous activation of Cl^- and K^+

* To whom correspondence should be addressed.

Fig. 1 Traces of whole-cell current from single acinar cells of mouse exorbital lacrimal glands. The acinar cells are voltage-clamped at a holding potential very close (within 3 mV) to the zero current potential, namely the resting membrane potential of the internally perfused cells. The internal solution contains 0.5 mM EGTA and no added calcium. From this holding potential a series of depolarizing and hyperpolarizing voltage jumps are superimposed (100–200 ms duration) evoking outward (upward deflections) and inward (downward deflections) currents respectively. Membrane potential (mV) indicates the holding potential and the membrane potential during the superimposed voltage jumps. *a*, Currents in the unstimulated cell and changes on exposing the cell to ACh, 10^{-5} M. ACh stimulation is associated with a sustained increase in the amplitude of the outward current and a largely transient increase in the inward currents, an effect blocked by atropine at 10^{-5} M in each of eight experiments (not shown). *b*, At the point indicated, the internal perfusion fluid has been changed to a solution 10^{-6} M in free calcium (*i*, intracellular). The rise in intracellular free calcium is associated with a sustained increase in the amplitude of the outward current (for each of six experiments). *c*, Effects of extracellular application of ACh (10^{-5} M), as in *a*. At the point indicated, the control intracellular solution is replaced by one containing 10 mM EGTA and no added calcium. Dialysis of the cell with this calcium chelator reduces the ACh-activated currents to pre-stimulus levels (in five experiments). *d*, The acinar cell is bathed in a calcium-free extracellular solution and at the point indicated is exposed to 10^{-5} M ACh. An initial increase in both outward and inward currents is not sustained in the absence of extracellular calcium (in each of six experiments). Re-introduction of extracellular calcium in the continued presence of the agonist (ACh) again gives a sustained increase in outward current amplitude (in each of four experiments). *e*, Removal of extracellular calcium has no effect on whole-cell currents in the unstimulated lacrimal acinar cell (in seven experiments). Stimulation with 10^{-5} M ACh is again associated with a transient increase in current amplitudes, and ACh removal after 4 min gives only a small further decrease in whole-cell currents. In the absence of ACh, the re-introduction of extracellular calcium produces no change in current amplitudes, but the cells respond upon readmission of ACh (in five experiments).

Methods. Individual acinar cells from exorbital lacrimal glands from adult male mice were isolated by incubation with trypsin and collagenase¹⁷. Standard patch clamp techniques were used to monitor voltage clamp currents from single cells using tight-seal, whole-cell recording mode¹⁹. The standard extracellular solution contained (mM): NaCl, 140; KCl, 4.7; CaCl₂, 1.2; MgCl₂, 1.13; glucose, 10; HEPES, 10, pH 7.2. All experiments were performed at 22–24 °C. Extracellular solutions without calcium had 1 mM EGTA added, and barium solutions had 88 mM BaCl₂ substituted for NaCl. The standard intracellular solution contained (mM): potassium glutamate, 80; KCl, 40; NaCl, 20; MgCl₂, 5; Na ATP, 4; glucose, 10; HEPES, 10; GTP, 2×10^{-5} M. The free calcium concentration was adjusted with EGTA or calcium/EGTA buffer in low concentration (≤ 0.5 mM)^{20,21}. Ins P₃ and Ins P₄ were included at 10^{-5} M. As sodium-proton exchange may be a prerequisite for Ca²⁺ mobilization in human platelets²² and an alkaline pH helps Ins P₃ release Ca²⁺ from permeabilized platelets²³, internal solutions were titrated to pH of 7.6. Ins P₃ and Ins P₄ were prepared as described²⁴. After formation of gigaohm seal, cells were moved into the outflows of superfusing tubes delivering extracellular solutions under gravity. Composition of the intracellular solution dialysing the cells was regulated by internal perfusion of the recording pipettes^{10,11,25}. The concentrations of K⁺ and Cl⁻ in the intracellular and extracellular solutions were constant. The equilibrium potentials for these ions are -80 mV and -20 mV respectively. The cells were voltage-clamped at a holding potential close (within 3 mV) to the zero current potential. Depolarizing (to 0 mV) and hyperpolarizing (to -80 mV) voltage jumps were superimposed from this level to evoke outward (upward) and inward (downward) currents. At 0 mV a large electrochemical gradient favours K⁺ current (difference between the K⁺ equilibrium Nernst potential and membrane potential is about 80 mV), and large noisy outward currents with short (<50 ms), relaxation time constants (see Fig. 3) are obtained, which are characteristic for high conductance Ca²⁺- and voltage-activated (maxi) K⁺ channels¹⁸. There is no evidence for other K⁺-selective channels from single channel current studies^{17,18,26}. At 0 mV membrane potential there is a small electrochemical gradient favouring outward Cl⁻ current (about 20 mV), but at [Ca²⁺]_i < 0.5 μ M, the Cl⁻ current is negligible. The inward Cl⁻ current at -80 mV is measured; this would equal, or be larger than, the outward Cl⁻ current at zero potential²⁸. At high [Ca²⁺]_i (1 μ M and above) the current-voltage relation for the whole-cell Cl⁻ current is linear²⁸ and the numerical value of the outward Cl⁻ current is only about one third of the inward Cl⁻ current at -80 mV (ref. 28). As the ACh-evoked increase in inward current at -80 mV is transient and smaller (even absent, as in Figs 2, 3 and 4) than the increased outward current at 0 mV, the latter must be dominated by K⁺ current. The former is abolished during Cl⁻-free conditions and it is therefore assumed to be carried by Cl⁻ (refs 26–28).

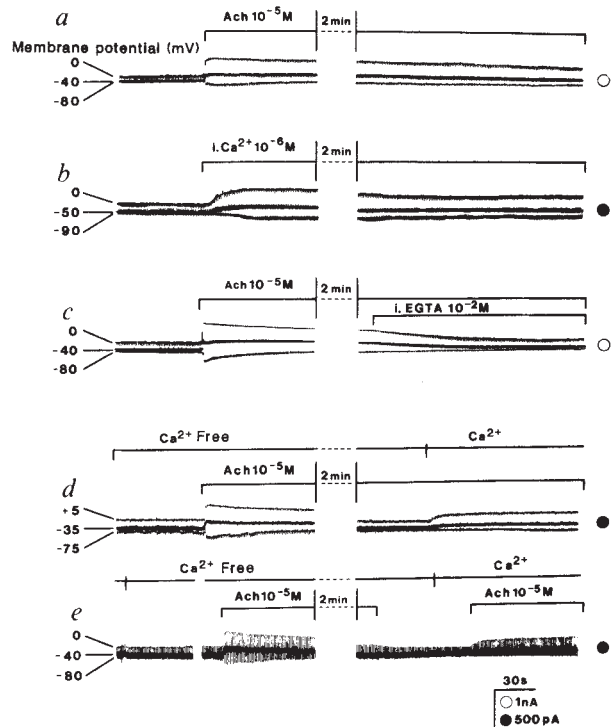
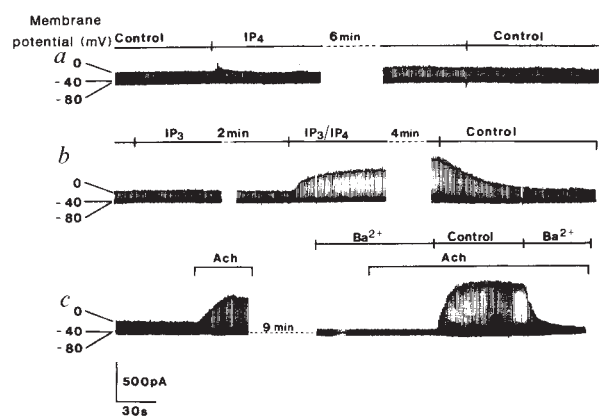


Fig. 2 Traces are sections of one continuous recording from the same lacrimal acinar cell. The intracellular solution had 0.5 mM EGTA with no added calcium. As for Fig. 1, the holding potential (-40 mV) is very close to the zero current potential and depolarizing and hyperpolarizing voltage jumps are superimposed (100 ms duration). *a*, The effect of changing the intracellular solution from the control to one containing 10^{-5} M Ins P₄ (IP₄). The Ins P₄ is present for ~8 min before return to control intracellular solution. *b*, Traces from left to right show the effect of 10^{-5} M Ins P₃ in the internal solution, which after ~4 min is replaced with a solution containing both Ins P₃ and Ins P₄. Only in the presence of the combined Ins P₃ and Ins P₄ is there any marked or sustained change in the amplitude of the outward currents (12 experiments). The increased amplitude of the outward current is sustained until Ins P₃/Ins P₄ is removed and the control intracellular solution reintroduced. *c*, Effect of ACh (10^{-5} M) in the extracellular solution (left), showing an increase in amplitude of the outward currents; effect of extracellular barium on whole-cell currents is shown to the right, when a marked reduction in the currents in the unstimulated cell is seen and the ACh-induced increase in outward current is abolished. Barium effects are reversible and reproducible in four experiments.



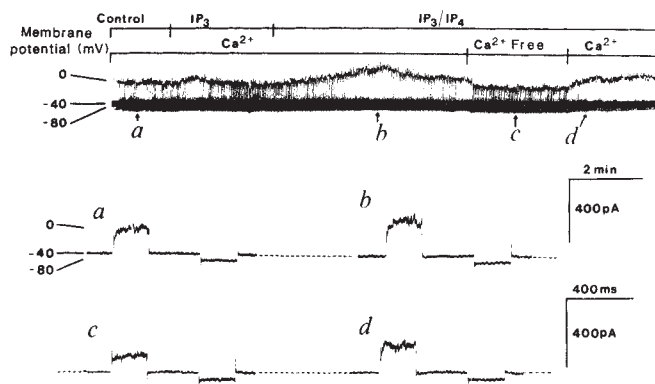


Fig. 3 Whole-cell current recording (top) showing the effect of introducing 10^{-5} M Ins P_3 (IP_3) into the internal perfusion fluid with $\sim 10^{-7}$ M Ca^{2+} (0.1 mM total EGTA). There is a transient increase in outward currents which return within the first minute to pre-stimulus levels. The subsequent introduction of Ins P_4 in the continued presence of Ins P_3 (IP_3/IP_4) gives an increase in outward current amplitude that is sustained even after 5 min. Removal of extracellular calcium then causes a fall in amplitude of outward currents to the level of the control intracellular solution (seven experiments). This reduction is reversed when extracellular calcium is reintroduced in the continued presence of Ins P_3 /Ins P_4 (four experiments). The oscilloscope traces *a-d* illustrate the time course and amplitude of the outward and inward currents at the points indicated on the continuous trace.

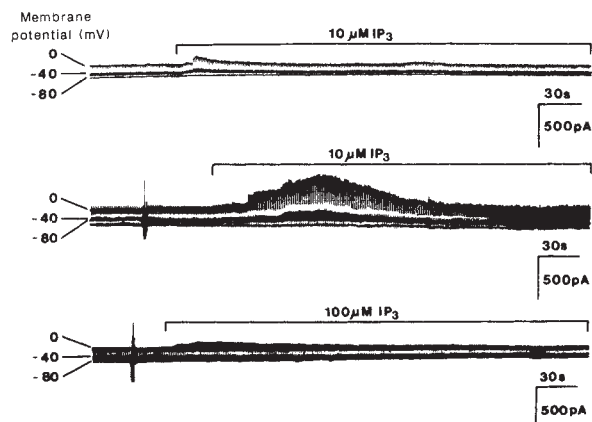


Fig. 4 Whole-cell current recordings illustrating the variable response of the lacrimal acinar cell to internal perfusion of Ins P_3 at 10^{-5} M (top two traces) and 10^{-4} M (lower trace). The internal solution contains Ca^{2+} at $\sim 10^{-7}$ M (0.1 mM total EGTA). Transient increases in K^+ current were observed in over 50% of cells (10 out of 18 experiments). The amplitude and duration of the transient increase in outward current was highly variable and showed no correlation with the concentration of Ins P_3 used. If Ins P_3 was included in an intracellular solution containing 0.5 mM EGTA with no added calcium, $[Ca^{2+}]$ was $\sim 10^{-8}$ M and Ins P_3 failed to evoke any increase in K^+ current at any of the concentrations tested: 10^{-5} M ($n=5$), 2×10^{-5} M ($n=2$) and 10^{-4} M ($n=2$).

channels^{33,34}, and this actually occurs in many of our whole-cell current experiments (Fig. 1). The Ca^{2+} -sensitivity of the Cl^- channels is much lower than that of the K^+ channels²⁶, and Fig. 1 shows that challenge with Ca^{2+} at the high concentration of 1 μ M only causes a very slow increase in the inward Cl^- current. We found the inward Cl^- current response, in agreement with a previous report²⁶, to be much more variable than the outward K^+ current response, and in some experiments there was no inward current response to ACh or to Ins P_3 /Ins P_4 (Figs 2 and 3).

A previous patch-clamp current study has demonstrated an external Ca^{2+} requirement for sustained K^+ channel-opening evoked by receptor (cholecystokinin) activation of pancreatic acinar cells³⁵. This Ca^{2+} requirement could be localized at a site separated from the hormone-receptor site, which indicated messenger-mediated rather than direct receptor-operated Ca^{2+} entry³⁵. Our results provide evidence that the messengers regulating Ca^{2+} entry in mammalian exocrine acinar cells, and probably

also in other cell types, are Ins P_3 and Ins P_4 acting together, as already proposed on the basis of injection studies on sea urchin egg cells⁷. There is evidence for the rapid generation of first Ins P_3 and then Ins P_4 in rat pancreatic acinar cells, following activation of muscarinic ACh receptors as well as cholecystokinin receptors, with a maximal increase in Ins P_4 concentration occurring within 15–30 seconds³⁶. We suggest that Ca^{2+} probably enters the cell through an Ins P_3 -sensitive Ca^{2+} store, as previously proposed^{38,39}, and that Ins P_4 modulates Ca^{2+} entry into that Ca^{2+} store³⁷. We should emphasize, however, that our data do not distinguish between the contributions made individually by the two inositol phosphates, and the only certain conclusion is that both are necessary to control Ca^{2+} entry in this experimental system. Several criteria for a second messenger function for Ins P_4 have been fulfilled and we may now have the key elements to understand the link between Ca^{2+} entry and the increased turnover of phospholipids after stimulation originally proposed by Michell⁴⁰.

Received 2 October; accepted 30 October 1987.

1. Streb, H., Irvine, R. F., Berridge, M. J. & Schulz, I. *Nature* **306**, 67–69 (1983).
2. Streb, H., Bayerdorffer, E., Haase, W., Irvine, R. F. & Schulz, I. *J. Membrane Biol.* **81**, 241–253 (1984).
3. Burgess, G. M. *et al.* *Nature* **309**, 63–66 (1984).
4. Prentki, M. *et al.* *Nature* **309**, 562–564 (1984).
5. Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315–321 (1984).
6. Irvine, R. F., Letcher, A. J., Heslop, J. P. & Berridge, M. J. *Nature* **320**, 631–634 (1986).
7. Irvine, R. F. & Moor, R. M. *Biochem. J.* **240**, 917–920 (1986).
8. Kuno, M. & Gardner, P. *Nature* **326**, 301–304 (1987).
9. Llano, I., Marty, A. & Tanguy, J. *Pflügers Arch. ges. Physiol.* **409**, 499–506 (1987).
10. Kimura, J., Noma, A. & Irisawa, H. *Nature* **319**, 596–597 (1986).
11. Jauch, P., Petersen, O. H. & Lauger, P. *J. Membrane Biol.* **94**, 99–115 (1986).
12. Petersen, O. H. & Maruyama, Y. *Nature* **307**, 693–696 (1984).
13. Petersen, O. H. & Findlay, I. *Physiol. Rev.* **67**, 1054–1116 (1987).
14. Iwatsuki, N. & Petersen, O. H. *J. Physiol., Lond.* **275**, 507–520 (1978).
15. Iwatsuki, N. & Petersen, O. H. *Pflügers Arch. ges. Physiol.* **377**, 185–187 (1978).
16. Pearson, G. T. & Petersen, O. H. *Pflügers Arch. ges. Physiol.* **400**, 51–59 (1984).
17. Findlay, I. *J. Physiol., Lond.* **350**, 179–195 (1984).
18. Trautmann, A. & Marty, A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 611–615 (1984).
19. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).

20. Findlay, I., Dunne, M. J. & Petersen, O. H. *J. Membrane Biol.* **83**, 169–175 (1985).
21. Iwatsuki, N. & Petersen, O. H. *Biochim. biophys. Acta* **819**, 249–257 (1985).
22. Siffert, W. & Akkerman, J. W. N. *Nature* **325**, 456–458 (1987).
23. Brass, L. F. & Joseph, S. K. *J. Biol. Chem.* **260**, 15172–15179 (1985).
24. Irvine, R. F., Letcher, A. J., Lander, D. J. & Berridge, M. J. *Biochem. J.* **240**, 301–304 (1986).
25. Soejima, M. & Noma, A. *Pflügers Arch. ges. Physiol.* **400**, 424–431 (1984).
26. Marty, A., Tan, Y. P. & Trautmann, A. *J. Physiol., Lond.* **357**, 293–325 (1984).
27. Findlay, I. & Petersen, O. H. *Pflügers Arch. ges. Physiol.* **403**, 328–330 (1985).
28. Evans, M. G. & Marty, A. *J. Physiol., Lond.* **378**, 437–460 (1986).
29. Maruyama, Y. & Petersen, O. H. *J. Membrane Biol.* **79**, 293–300 (1984).
30. Connolly, T., Bansal, V. S., Bross, T. E., Irvine, R. F. & Majerus, P. W. *J. Biol. Chem.* **262**, 2146–2149 (1987).
31. Biden, T. J. & Wollheim, C. B. *J. Biol. Chem.* **261**, 11931–11934 (1986).
32. Zilberman, Y., Howe, L. R., Moore, J. P., Heskeith, T. R. & Metcalf, J. C. *EMBO J.* **6**, 957–962 (1987).
33. Petersen, O. H. *Am. J. Physiol.* **251**, G1–G13 (1986).
34. Petersen, O. H. & Gallacher, D. V. *A. Rev. Physiol.* **50**, 65–80 (1988).
35. Suzuki, K., Petersen, C. C. H. & Petersen, O. H. *FEBS Lett.* **192**, 307–312 (1985).
36. Trimble, E. R., Bruzzone, R., Meehan, C. J. & Biden, T. J. *Biochem. J.* **242**, 289–292 (1987).
37. Irvine, R. F. & Moor, R. M. *Biochem. biophys. Res. Commun.* **146**, 284–290 (1987).
38. Putney, J. W. Jr *Cell Calcium* **7**, 1–12 (1986).
39. Putney, J. W. Jr *Am. J. Physiol.* **252**, G149–G157 (1987).
40. Michell, R. H. *Biochim. biophys. Acta* **415**, 81–147 (1975).

Occurrence and extracellular actions of inositol pentakis- and hexakisphosphate in mammalian brain

Mario Vallejo*, Trevor Jackson*, Stafford Lightman† & Michael R. Hanley*

* MRC Molecular Neurobiology Unit, University of Cambridge Medical School, Hills Road, Cambridge CB2 2QH, UK

† Medical Unit, Charing Cross and Westminster Medical School, Page Street Wing, Westminster Hospital, London SW1 2AP, UK

Although inositol 1,3,4,5,6-pentakisphosphate (InsP₅) and hexakisphosphate (InsP₆) have been recognized for some time as naturally-occurring metabolites of inositol¹, their occurrence in mammalian cell types²⁻⁵, including one of neural origin⁶, has only recently been documented. This is of interest because of the recognized second messenger role of inositol 1,4,5-trisphosphate (InsP₃) in intracellular signalling; coupling surface stimuli to cytoplasmic calcium discharge⁷. The metabolism, existence in normal mature tissues, and possible functional roles of these inositol polyphosphates are unknown. Here we report evidence that InsP₅ and InsP₆ are synthesized in intact brain after labelling with [³H]inositol *in vivo*. We also show that local infusion of InsP₅ and InsP₆ into a discrete brain stem nucleus implicated in cardiovascular regulation, results in dose-dependent changes in heart rate and blood pressure.

Sprague-Dawley rats were anaesthetized with equithesin, and [³H]myo-inositol was injected stereotactically into the third ventricle and the rats allowed to recover. Previous studies (Moratalla, R., M.V. and S.L., in preparation) have shown that [³H]myo-inositol is incorporated into brain lipids in a time-dependent way, reaching plateau levels 24 h after intracerebroventricular administration. At 24 h, the animals were decapitated and their brains rapidly removed and dissected for subsequent determination of the regional content of radiolabelled inositol phosphates by ion-exchange high-pressure liquid chromatography (HPLC) as previously described⁶. Incorporation of [³H]myo-inositol into membrane lipids was also determined and used to normalize the concentration of inositol phosphates as a percentage of labelled lipids.

Chromatographic analysis of these samples revealed an elution pattern qualitatively similar to that seen in mammalian cell lines *in vitro*^{2-4,6}, with the exception that no radioactivity was seen in the position of the Ins1,3,4,5P₄ standard (Fig. 1a). Two labelled peaks with similar chromatographic properties to InsP₅ and InsP₆ (ref. 2) were found at variable levels in different regions of the rat brain (Fig. 1b). We have taken the precaution of identifying these peaks as 'InsP₅' and 'InsP₆', rather than specific positional- or stereoisomers, as their precise structural identification is incomplete. In chromatography, however, InsP₅ peaks, reproducibly and exactly, with a rigorously characterized ³²P-labelled myo-inositol-1,3,4,5,6-pentakisphosphate standard from avian erythrocytes (Fig. 1a legend). These levels cannot be assumed to correspond to authentic concentrations, but they do imply that the levels of InsP₅ and InsP₆ are independently regulated. Both ³H-labelled InsP₅ and InsP₆ were found in about the same proportions in the midbrain and hypothalamus. InsP₅, however, was more abundant than InsP₆ in the medulla oblongata, whereas the opposite was seen in the hippocampus—the region with the greatest overall concentration of InsP₆. No labelled InsP₅ could be detected in the striatum, although labelled InsP₆ was present. In general, the brain levels of labelled InsP₅ and InsP₆ are much lower than that of labelled inositol monophosphates, which range from 15 to 25% of lipid content.

Although the specific pathways through which myo-inositol is converted to InsP₅ and InsP₆ are not yet established, our data

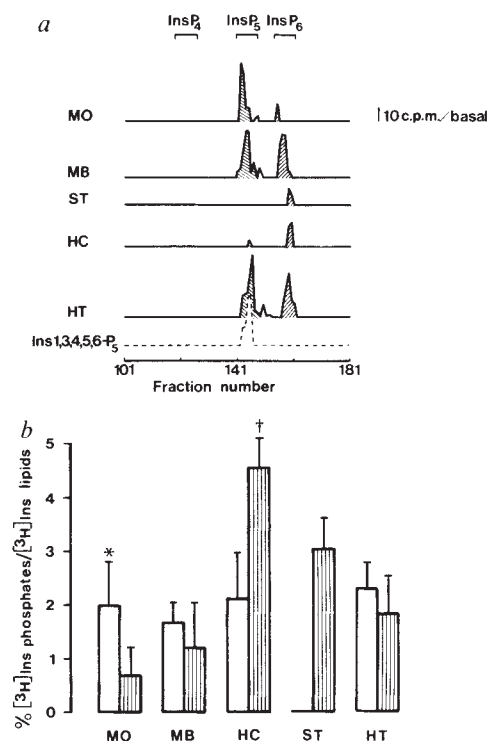
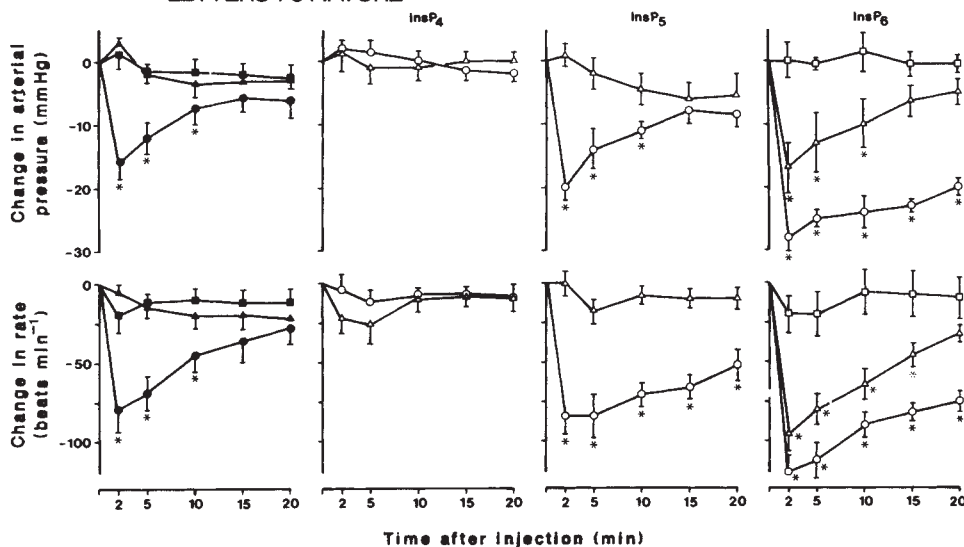


Fig. 1 a, Representative elution profiles of inositol phosphates from InsP₄ to InsP₆. Samples from different brain regions were analysed by ion-exchange HPLC (described below). Peak identification was carried out by co-chromatography with ³²P-labelled standards. The dotted line shows an internal standard of authentic [³²P] Ins(1,3,4,5,6) P₅ added to the hypothalamus sample before chromatography. It is exactly coincident with the major cross-hatched peak identified as 'InsP₅'. The elution positions of these standards are not absolute, due to the very high salt elution, but do not vary more than three fractions. The bars show the standard positions which were confirmed by use of a ³²P-labelled internal standard. The minor peaks of ³H-labelled material may correspond to other inositol-pentakisphosphate isomers but are not always observed. b, Normalized levels of measured inositol phosphates relative to labelled inositol lipids. Inositol phosphates were determined as the total integrated area under the peak (shown by cross-hatching in a) with an integration threshold of 30 c.p.m. (average background = 19 c.p.m.), corrected to a percentage of total [³H]inositol-labelled lipid in the same sample¹⁵. The results shown are means ± s.e.m. for 4–7 individual determinations from different microinjected animals. MO, medulla oblongata; MB, midbrain; HC, hippocampus; ST, striatum; HT, hypothalamus. * *P* < 0.05, † *P* < 0.02 (Student's *t*-test, InsP₅ versus InsP₆). Open bars, InsP₅; shaded bar, InsP₆.

Methods. Male Sprague-Dawley rats (270–330 g) were anaesthetized with equithesin (4.5 ml kg⁻¹ intraperitoneally (i.p.) and placed in a Kopf stereotaxic frame with the incisor bar 5 mm above the level of the ear bars. D-my-[2,3-³H]inositol (6 μCi) (specific activity 54.5 Ci mmol⁻¹) (New England Nuclear) were injected in the third ventricle in a volume of 10 μl over 2 min, using a 10 μl Hamilton syringe equipped with a 28 gauge needle. This formed a lateral angle of 10° with the vertical axis to avoid damage of the sagittal sinus. Stereotaxic coordinates used were: RC, 0.2 and L, 1.4 mm from bregma; DV 7.9 mm from the dorsal surface of the dura. Correct placement of the injectate using these coordinates was examined in a separate group of rats by administering methylene blue (10 μl) and checking its distribution through the ventricular system. The animals were allowed to recover and then decapitated 24 h after the injection. The brains were quickly removed and dissected on ice¹⁶. Each sample was then sonicated in 10% (w/v) perchloric acid and kept at -20 °C until assayed. After centrifugation the supernatant was removed and neutralized with a 1:1 mixture of 1,1,2-trichloro-trifluoroethane: tri-*n*-octylamine. After a further centrifugation the inositol-phosphate-containing upper phase was subjected to ion-exchange HPLC on a Partisil SAX 10 μm column as described⁶. Before sample injection, ³²P-labelled standards (200–300 c.p.m.) were added and counted by a ³H/³²P dual-label program. Fractions of 0.25 ml were collected and radioactivity determined by liquid-scintillation counting. Pellets were washed with 1 mM EDTA and inositol lipids were extracted with 1.88 ml chloroform:methanol (1:2), containing 0.1 M HCl. This was followed by a further 0.62 ml chloroform and 0.62 ml H₂O. After mixing and centrifugation the upper phase was removed and replaced by chloroform:methanol:0.1 M HCl (3:47:48). This was again mixed and centrifuged, after which the lower phase was removed, dried down and [³H]inositol-lipid content determined by liquid-scintillation counting. Chromatographic recoveries of water soluble and organic-solvent extracts were in all cases >90%.

Fig. 2 Changes in mean arterial pressure and heart rate seen after microinjections of saline, glutamate or inositol phosphates into the nucleus tractus solitarius of rats. Each point represents the mean $\pm$ s.e.m. of 4–7 injections * $P < 0.05$, Duncan's multiple-range test, as compared with the saline-treated group.

Methods. Male Sprague-Dawley rats (250–350 g) were anaesthetized with urethane (1.3 g kg^{-1} i.p.). The left common carotid artery was cannulated with polythene tubing and connected via a Statham P23 ID transducer to a Gould thermal chart recorder for continuous measurement of blood pressure and heart rate. The animals were placed in a Kopf stereotaxic frame with the head flexed downwards to an angle of 45° . After a midline incision through the skin, the dorsal neck muscles were retracted and the atlanto-occipital membrane was incised to expose the dorsal surface of the brainstem. After 20–30 min, an injection ($0.2 \mu\text{l}$ delivered over 10 s) was administered into the right nucleus tractus solitarius, using pulled glass micropipettes ($50 \mu\text{m}$ outer diameter) connected by plastic tubing to a $10 \mu\text{l}$ Hamilton syringe driven by a microinfusion pump. The injections were placed at the level of the caudal tip of the area postrema, 0.3 mm lateral to the midline and 0.6 mm below the dorsal surface of the brainstem. Rats were injected with vehicle (0.9% saline), L-glutamic acid, sodium salt (BDH), inositol hexakisphosphate, sodium salt (Sigma), inositol 1,3,4,5,6-pentakisphosphate, (sodium salt, converted from barium salt (Cal Biochem) by ion exchange¹) or mixture of inositol-tetrakisphosphate isomers prepared by acid hydrolysis of InsP_6 (ref. 1). Mixed InsP_4 isomers were purified by ion exchange on Dowex AG1X8 (formate form)⁶, and desalted as described⁵. Inositol phosphate concentrations were determined by quantitative phosphate analysis¹⁷. Each rat received only one injection. To confirm correct placement, histological examinations of frozen sections of the brains were carried out in rats injected with methylene blue into the same positions. Sections were cut at $25 \mu\text{m}$, stained with thionine and compared to the atlas of Pellegrino *et al.*¹⁸, revealing that the injectate was distributed within the nucleus-tractus-solitarius area, with minimal diffusion into the adjacent dorsal motor nucleus of the vagus. Anatomical specificity was further tested by injecting 200 pmol InsP_5 ($n = 3$) or InsP_6 ($n = 2$) 1 mm lateral to the nucleus tractus solitarius and 1 mm below the surface of the brainstem. No cardiovascular changes were elicited by these injections. Controls: ■, saline; ▲, Glu 2 nmol ; ●, Glu 5 nmol . Ins phosphates: □, 20 pmol ; △, 40 pmol ; ○, 200 pmol .



show that these inositol phosphates are not only synthesized in normal mammalian brain, but that the synthesis is regionally regulated, perhaps reflecting different functional requirements.

There is no evidence yet for the involvement of these inositol phosphates in intracellular signalling mechanisms. In stimulated cells they do not exhibit the transient alterations in levels observed in, for example, the InsP_4 and InsP_3 isomers^{3–6}. Indeed, their detection in culture relies on prolonged incubation with [^3H]inositol, suggesting they may be part of a different metabolic pool. One possibility is that the inositol polyphosphates may act as extracellular signals. To test this possibility, we investigated the *in vivo* response to microinjections of exogenous InsP_5 and InsP_6 into the nucleus tractus solitarius (NTS). This nucleus, located in the dorsomedial medulla oblongata, is involved in the control of cardiovascular homeostasis and is the site where primary baroreceptor afferents first synapse⁸. It contains a number of neuroactive substances, some of which, after local infusion, induce rapid and specific changes in cardiovascular parameters^{9–11}. The synthesis of InsP_5 and InsP_6 in the brainstem suggested that the NTS could be used to evaluate possible extracellular effects.

Unilateral microinjections were made into the right NTS of urethane-anaesthetized Sprague-Dawley rats with a glass micropipette after stereotaxic surgery. Basal mean arterial pressure and heart rate were $89.2 \pm 3.3 \text{ mm Hg}$ and $391 \pm 8 \text{ beats min}^{-1}$, respectively. Microinjection of saline (0.9% NaCl) or 40 pmol InsP_5 had no significant effect, but higher doses (200 pmol) resulted in a marked decrease in both mean arterial pressure and heart rate (Fig. 2). This hypotensive effect developed rapidly and peaked within 2–5 min, after which a slow recovery in both parameters was seen.

Microinjections of InsP_6 into the NTS also produced marked effects, but at lower doses than InsP_5 . Administration of 40 pmol InsP_6 resulted in hypotension and bradycardia, similar to that observed with 200 pmol InsP_5 (Fig. 2). Microinjections of 200 pmol InsP_6 produced an even bigger fall in both blood

pressure and heart rate which also developed rapidly, peaking after 2 min (Fig. 2). For comparison, microinjections of the excitatory amino acid glutamic acid (2 nmol) into the NTS (see for example ref. 12) did not produce any significant alteration of cardiovascular parameters, but caused hypotension and bradycardia at a higher dose (5 nmol) (Fig. 2). InsP_5 and InsP_6 are the most potent small molecules known to induce this response, being 10–100 $\times$ more potent than glutamic acid.

To test the structural specificity of the cardiovascular responses, the NTS was microinjected with equimolar doses of a mixture of InsP_4 isomers. When administered at doses of 40 , 200 (Fig. 2) or 400 pmol (data not shown), no significant cardiovascular changes were seen. Nonspecific effects due to common chemical properties of inositol phosphates, are therefore unlikely. The observed responses appear to result exclusively from a central mechanism, as intravenous administration of up to 600 pmol InsP_5 ($n = 3$) or InsP_6 ($n = 3$) produced no haemodynamic changes. Anatomical specificity was also tested by injection adjacent to the NTS (see Fig. 2, legend), and the lack of cardiovascular response suggests that both inositol polyphosphates act on cells within the NTS.

The detection of InsP_5 and InsP_6 , but not any isomer of InsP_4 , in normal and fully differentiated mammalian brain tissue suggests that inositol polyphosphates are not exclusively associated with cell lines. This must be taken into account when considering the possible functional significance of these substances. The inability to detect significant levels of any InsP_4 isomer does not imply that none is produced in normal brain, but rather that the experimental design, metabolic labelling without external stimulation, favours detection of inositol metabolites that accumulate, rather than those that are transiently produced. Thus, the regional specificity of the pattern of InsP_5 and InsP_6 levels suggests independent metabolism rather than a strict precursor-product relationship between the two isomers. The identification of the conversion pathways and intermediates in the synthesis and breakdown of the InsP_5 and InsP_6 will be of great interest, and may show tissue-specific variations. Indeed,

a specific InsP_4 kinase has been shown in several tissues and appears to be enriched in brain¹³.

Both InsP_5 and InsP_6 are found in the brainstem and strongly stimulate systemic haemodynamic effects when locally administered to this region. These effects are reversible, dose dependent, and anatomically and structurally specific, whereas InsP_4 isomers are inactive. The potency and specificity of these effects suggests that inositol polyphosphates could be novel pharmacological tools, particularly for the investigation of the central control of blood pressure. These pharmacological effects also predict a physiological role, which may be an example of the nervous system specifically compartmentalizing a widespread small molecule, such as an amino acid or a nucleotide¹⁴, to act as an intercellular signal. This possibility is strongly supported by the finding that InsP_5 and InsP_6 , but not InsP_4 , specifically and reversibly excite regional populations of neurons.

We thank Drs R. Irvine, C. P. Downes, L. Stephens and P. T. Hawkins for critical discussions, advice and for the inositol-phosphate standards. M.V. was supported by an EMBO long-term Fellowship. T.J. was supported by an MRC Partnership Award with S.K. & F. Research. M.H. is the recipient of a Research Award from the International Life Sciences Research Institute.

Received 13 July; accepted 5 November 1987.

1. Cosgrove, D. J. *Ann. N.Y. Acad. Sci.* **165**, 677-686 (1969).
2. Heslop, J. P., Irvine, R. F., Tashjian, A. H. Jr. & Berridge, M. J. *J. exp. Biol.* **119**, 395-401 (1985).
3. Heslop, J. P., Blakely, D. M., Brown, K. D., Irvine, R. F. & Berridge, M. J. *Cell* **47**, 703-709 (1986).
4. Tilly, B. C. *et al. Biochem. J.* **244**, 129-135 (1987).
5. Hawkins, P. T., Stephens, L. & Downes, C. P. *Biochem. J.* **238**, 507-516 (1986).
6. Jackson, T. R., Hallam, T. J., Downes, C. P. & Hanley, M. R. *EMBO J.* **6**, 49-54 (1987).
7. Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315-321 (1984).
8. Palkovits, M. & Zaborszky, L. *Prog. Brain Res.* **47**, 9-34 (1977).
9. DiMicco, J. A., Gale, K., Hamilton, B. & Gillis, R. A. *Science* **204**, 1106-1108 (1979).
10. Carter, D. A., Vallejo, M. & Lightman, S. L. *Peptides* **6**, 421-425 (1985).
11. Vallejo, M. & Lightman, S. L. *Brain Res.* **422**, 295-302 (1987).
12. Talman, W. T., Perrone, M. H. & Reis, D. J. *Science* **209**, 813-815 (1980).
13. Stephens, L., Hawkins, P., Morris, A. & Downes, C. P. *Biochem. J.* (in the press).
14. Cooper, J. R., Bloom, F. E. & Roth, R. H. *The Biochemical Basis of Neuropsycharmacology* 3rd edn (Oxford University Press, London, 1984).
15. Sugiyama, H., Tennes, K. A. & Putney, J. W. *Biochem. J.* **244**, 647-653 (1987).
16. Glowinski, J. & Iversen, L. L. *J. Neurochem.* **13**, 655-669 (1966).
17. Bartlett, G. R. *J. Biol. Chem.* **234**, 466-468 (1959).
18. Pellegrino, L. J., Pellegrino, A. S. & Cushman, A. J. *A Stereotaxic Atlas of the Rat Brain* (Plenum, New York, 1979).

Interleukin-1 regulates synthesis of nerve growth factor in non-neuronal cells of rat sciatic nerve

Dan Lindholm, Rolf Heumann, Michael Meyer & Hans Thoenen

Department of Neurochemistry, Max-Planck-Institute for Psychiatry, D-8033 Planegg-Martinsried, FRG

The Schwann cells and fibroblast-like cells of the intact sciatic nerve of adult rats synthesize very little nerve growth factor (NGF) (ref. 1). After lesion, however, there is a dramatic increase in the amounts of both NGF-mRNA and NGF protein synthesized by the sciatic non-neuronal cells^{1,2}. This local increase in NGF synthesis partially replaces the interrupted NGF supply from the periphery to the NGF-responsive sensory and sympathetic neurons, whose axons run within the sciatic nerve¹. Macrophages, known to invade the site of nerve lesion during wallerian degeneration^{3,4}, are important in the regulation of NGF synthesis⁵. Here we demonstrate that the effect of macrophages on NGF-mRNA levels in cultured explants of sciatic nerve can be mimicked by conditioned media of activated macrophages, and that interleukin-1 is the responsible agent.

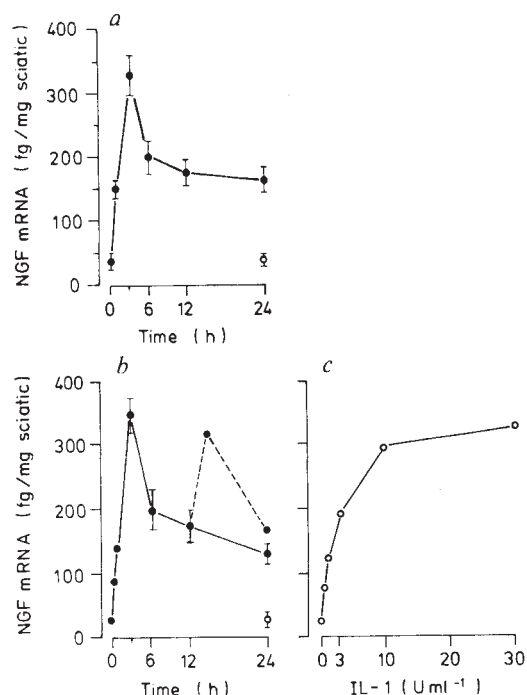


Fig. 1 *a*, Changes in NGF-mRNA levels of rat sciatic nerve segments by conditioned media of activated rat macrophages. *b*, Time course of induction by IL-1β. Human recombinant IL-1β (30 U ml⁻¹) of specific activity 4 × 10⁴ U μg⁻¹ protein was used. The dashed line shows the effect of re-addition of IL-1β to cultures incubated 12 h with the protein. *c*, IL-1β dose-response curve for NGF-mRNA induction. The incubation was for three hours and averages of two determinations at each concentration are given. **Methods.** Sciatic nerve segments (3 cm) of adult Wistar rats were cultured in 1 ml Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. After three days culture, medium conditioned by either rat peritoneal macrophages or human recombinant IL-1β was added. Rat peritoneal macrophages were purified by attachment to plastic, and activated with lipopolysaccharide (LPS) (Sigma) for 20 h (70 μg ml⁻¹). Media conditioned by the adherent cells with more than 95% macrophages, as evaluated by the specific antibody ED1 (ref. 17), were collected and added to the nerve explant cultures. NGF-mRNA was determined by a quantitative Northern blot procedure, using a calibration standard (0.92 kilobases (kb)) for NGF-mRNA (ref. 16). The recovery of RNA was estimated using a shorter NGF-mRNA transcript (0.51 kb), which was added to the tissue samples before RNA extraction¹⁶. The autoradiograms were quantified on a scanning densitometer for the amount of NGF-mRNA present. Values are expressed as fg per mg wet weight of tissue, and they represent mean ± s.d. of three or more separate determinations. The open circles represent the control values of nerve segments incubated for 24 h.

In previous experiments we have shown that although the lesion-mediated increase in NGF-mRNA is biphasic *in vivo*¹, in cultures of rat sciatic nerve there was only a rapid and transient increase in NGF-mRNA, which was not followed by the chronic rise observed *in vivo*⁵. The addition of activated macrophages to the culture system resulted in increase of sciatic NGF-mRNA comparable to that observed *in vivo*⁵. To decide whether these effects result from cell-cell contact or are mediated by a humoral factor, we prepared media conditioned by activated rat macrophages. Figure 1 shows that rat sciatic nerve segments cultured for three days contained only relatively low amounts of NGF-mRNA which were markedly (11-fold) enhanced when the cultures were supplemented by macrophage-conditioned media, to reach a stable maximum after three hours. Of the many biologically active molecules synthesized by activated macrophages⁶, prostaglandin E₂, and acidic and basic fibroblast growth factor (FGF) did not affect the amount of NGF-mRNA

Table 1 Effect of different agents on rat sciatic nerve NGF-mRNA transcription

	Sciatic NGF-mRNA (fg per mg tissue)
Controls	34.5
+PGE ₂	47.2
+acidic FGF	28.0
+basic FGF	29.0
+PDGF	71.3
+TNF	75.2

Agents were added to sciatic nerve segments after culture for three days. After incubation for three hours, sciatic NGF-mRNA was quantitated as described in Fig. 1 legend. Prostaglandin E₂ (Sigma) was used at a final concentration of 5 μ M (ref. 18); acidic and basic FGF (a gift from Dr Sensenbrenner¹⁹) were at a concentration of 5 ng ml⁻¹. PDGF (Speywood, UK) was used at a concentration of 3 U ml⁻¹ (ref. 20). Human TNF- α (cachectin; Amgen, Amersham) was used at a final concentration of 1,000 U ml⁻¹ (ref. 21). These concentrations are comparable to those known to produce a maximal effect. Values represent the average of two or more separate determinations.

produced (Table 1). Tumour necrosis factor (TNF) and platelet-derived growth factor (PDGF), which are major constituents of macrophage-conditioned media⁷, doubled the amount of NGF-mRNA present. But the addition of recombinant human interleukin-1 β (IL-1 β) (ref. 8) resulted in a 14-fold increase in NGF-mRNA (Fig. 1b). The effect of IL-1 β was discernible at a concentration of 1 U ml⁻¹ and was maximal at about 10 U ml⁻¹. No difference was seen between the effects of recombinant IL-1 α and IL-1 β which are both produced by activated macrophages⁹, supporting the view that these structurally different molecules (the homology between human and murine IL-1 α and IL-1 β is only 30%¹⁰⁻¹²) have similar biological activities. The effects of human IL-1 and rat macrophage-conditioned medium on NGF-mRNA are similar, both with respect to the extent and the time-course of the increase. Re-addition of IL-1 β after 12 hours resulted in a repeat increase in NGF-mRNA to the level observed after the first addition, indicating that the protein was either not stable, or was degraded by the sciatic explants.

Because antibodies to rat IL-1 were not available, we used conditioned media of human macrophages, which affected the amounts of rat sciatic nerve NGF-mRNA produced to the same extent as the rat medium. The increase in NGF-mRNA observed after addition of recombinant IL-1 and human macrophage-derived medium could both be inhibited to the same extent by specific polyclonal antibodies to human IL-1 (Fig. 2). Thus, most, if not all, of the activity of the human macrophage-conditioned medium is due to IL-1. It is probable that the same is true for the conditioned medium of rat macrophages, although direct proof is not yet possible due to the unavailability of antibodies to rat IL-1 and to the poor interspecies immunological cross-reactivity¹³. Our interpretation is supported by the association of the activity of the rat-conditioned medium with fractions of relative molecular mass (M_r) between 10,000-20,000 (10-20K) (data not shown).

These results could explain the *in vivo* and the *in vitro* discrepancy⁵: *in vitro*, macrophages must be added to simulate the chronic increase in NGF-mRNA which occurs after lesion *in vivo*, and which remains high for more than two weeks^{1,5}. Following a peripheral nerve injury, macrophages invade the site of lesion^{3,4} to stimulate concomitant increase in NGF-mRNA levels⁵. These observations combined with ours indicate that macrophages are responsible for the persistent increase in NGF in the injured nerve *in vivo*. Note that macrophages themselves synthesize very little NGF (ref. 5), therefore their contribution

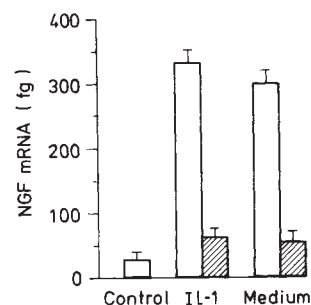


Fig. 2 Effect of macrophage-conditioned medium on sciatic NGF-mRNA; blockade by antibodies to IL-1. Conditioned media derived from LPS-activated human peritoneal macrophages were diluted 1 to 20 and used in the rat nerve explant cultures (see Fig. 1). A part of the medium was pretreated with rabbit antibodies raised against human IL-1 before adding it to the culture system. As a control 30 U of human recombinant IL-1 was treated similarly and then added to explant cultures. The antibodies used were polyvalent (Genzyme, Boston). Following incubation for three hours, the amounts of sciatic NGF-mRNA present were determined as described in Fig. 1 legend. The open bars show the results of incubating the nerve pieces with IL-1 or with the human macrophage-conditioned medium; the hatched bars represent results after pretreatment of the medium or of IL-1 with the specific antibodies. The values represent mean $\pm$ s.d. of three independent determinations.

to the amount of NGF present is indirect and mediated mainly by IL-1.

Macrophages and the IL-1 they release could also be important in the regulation of NGF synthesis during normal development; the amount of NGF-mRNA in the sciatic nerve of newborn rats is comparable to that in lesioned adult nerves⁵, and decreases to adult values by the third postnatal week⁵, by which time the high density of macrophages in the newborn rat sciatic nerve has also decreased to the low intact adult nerve levels⁴. Interestingly, IL-1 activity has also been detected in the central nervous system following brain injury¹⁴, but cellular site of synthesis has not yet been established^{13,15}.

We thank Dr A. Shaw (Biogen) and A. Gronenborn (Martinsried) for a gift of human recombinant IL-1. We are also grateful to Dr M. Blumenstein for help with obtaining human macrophages, to Elfi Grossmann for typing the manuscript, and to Drs D. Edgar and Y.-A. Barde for critically reviewing the manuscript. D.L. is a recipient of a fellowship from EMBO.

Received 17 August; accepted 20 October 1987.

- Heumann, R., Korsching, S., Bandtlow, C. & Thoenen, H. *J. cell Biol.* **104**, 1623-1631 (1987).
- Bandtlow, C. E., Heumann, R., Schwab, M. E. & Thoenen, H. *EMBO J.* **6**, 891-899 (1987).
- Perry, V. H., Brown, M. C. & Gordon, S. *J. exp. Med.* **165**, 1218-1223 (1987).
- Stoll, G. & Müller, H. W. *Neurosci. Lett.* **72**, 233-238 (1986).
- Heumann, R. *J. exp. Biol.* **132**, 133-150 (1987).
- Nathan, C. F., Murray, H. W. & Cohn, Z. A. *New Engl. J. Med.* **303**, 622-626 (1980).
- Shimokado, K. *et al. Cell* **43**, 277-286 (1985).
- Wingfield, P. *et al. Eur. J. Biochem.* **160**, 491-497 (1986).
- Oppenheim, J. J., Kovacs, E. J., Matsushima, K. & Durum, S. K. *Immun. Today* **7**, 45-56 (1986).
- March, C. J. *et al. Nature* **315**, 641-647 (1985).
- Lomedico, P. T. *et al. Nature* **312**, 458-461 (1984).
- Auron, P. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 7907-7911 (1984).
- Giulian, D., Baker, T. J., Shih, L.-C. N. & Lachman, L. B. *J. exp. Med.* **164**, 594-604 (1986).
- Giulian, D. & Lachman, L. B. *Science* **228**, 497-499 (1985).
- Fontana, A., Kristensen, F., Dubs, R., Gemsa, D. & Weber, E. *J. Immun.* **129**, 2413-2419 (1982).
- Heumann, R. & Thoenen, H. *J. Biol. Chem.* **261**, 9246-9249 (1986).
- Dijkstra, C. D., Dopp, E. A., Joling, P. & Kraal, G. *Immunology* **54**, 589-599 (1985).
- Brandwein, S. R. *J. Biol. Chem.* **261**, 8624-8632 (1986).
- Pettman, B., Labourette, G., Weibel, M. & Sensenbrenner, M. *Neurosci. Lett.* **68**, 175-180 (1986).
- Grennberg, M. E. & Ziff, E. B. *Nature* **311**, 433-438 (1984).
- Bevilacqua, M. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 4533-4537 (1986).

Specificity of peptide presentation by a set of hybrid mouse class I MHC molecules

Janet L. Maryanski*, Jean-Pierre Abastado†
& Philippe Kourilsky†

* Ludwig Institute for Cancer Research, Lausanne branch, Chemin des Boveresses, CH-1066 Epalinges, Switzerland
† Unité de Biologie Moléculaire, Unité 277 Institut National de la Santé et de la Recherche Médicale, Unité Associée 535 Centre National de la Recherche Scientifique, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

The class I molecules of the major histocompatibility complex (*H-2* in mouse, HLA in man) are membrane proteins composed of a polymorphic heavy chain associated with β -2-microglobulin. Recent studies suggest that class I molecules present peptides derived from processed antigens to the receptor of cytolytic T cells^{1,2}. In particular, in the *H-2*^d haplotype, synthetic HLA peptides can be recognized on *K*^d-bearing target cells by *K*^d-restricted cytolytic T cells specific for HLA (ref. 2). Here we analyse the specificity of presentation of two HLA peptides by a set of chimaeric *K*^d/*D*^d molecules³ to four different cytolytic T-cell clones. We identify two distinct regions within the second external (α 2) domain of *K*^d that contribute to its specificity as a restriction element. Our results indicate that the binding of an immunogenic peptide by a class I molecule is not always sufficient for its recognition by the T-cell antigen receptor. This suggests that the major histocompatibility complex restriction element either interacts with the T-cell antigen receptor or induces the recognized conformation of the peptide.

The cytolytic T-cell (CTL) clones used in this study were isolated from DBA/2 mice immunized with syngeneic (*H-2*^d) P815 cells transfected with cloned genes for HLA-CW3 or HLA-A24 (refs 2, 4 and 24). CTL clones CW3/10.1 and CW3/701.1 lyse P815-CW3 but not P815-A24 transfectant target cells, whereas CTL-A24/10.1 lyses P815-A24 but not P815-CW3 targets, and CTL-CW3/1.1 lyses both. In contrast to P815 cells, L cells (*H-2*^k) transfected with the same HLA genes were not lysed, except on introduction of a *K*^d gene⁴. These and other experiments^{4,5} indicated that HLA was recognized as a nominal antigen in the context of a *K*^d restriction element. It was further shown that the HLA antigen could be mimicked with synthetic peptides corresponding to the C-terminal end of the α 2 domain of HLA-CW3 or HLA-A24 (refs 2, 6, 24 and 25). Recognition of peptides CW3 170–182 and A24 170–182 that differ by only one amino-acid residue (see Fig. 1) corresponded with the pattern of lysis of P815-CW3 and P815-A24 transfectant target cells. Moreover, preliminary experiments indicated that the CW3 and A24 peptides were also specifically recognized on L cells transfected with *K*^d, but not with *D*^d genes (for example, see the controls in Fig. 2). It thus seemed feasible to analyse L cells transfected with *K*^d/*D*^d hybrid genes to map critical *K*^d-specific sites on the restriction element.

A large set of chimaeric *K*^d/*D*^d sequences that correspond to *K*^d at the 5'-end and to *D*^d at the 3'-end have been produced by *in vivo* recombination in *Escherichia coli*³. Eleven such hybrid complementary DNAs, termed R1 to R11, (Fig. 1) were engineered into a eukaryotic expression vector containing the SV40 early promoter and transfected into L cells. Transfected lines expressing reasonably high and comparable amounts of hybrid molecules were isolated by fluorescence-activated cell sorting⁷. Recombinant R6 was only expressed feebly⁷ and was therefore excluded from the present study.

These transfectants were analysed in the presence of HLA peptides CW3 170–182 and A24 170–182 for lysis by four different CTL clones (Fig. 2). Transfectants R10 and R11 that

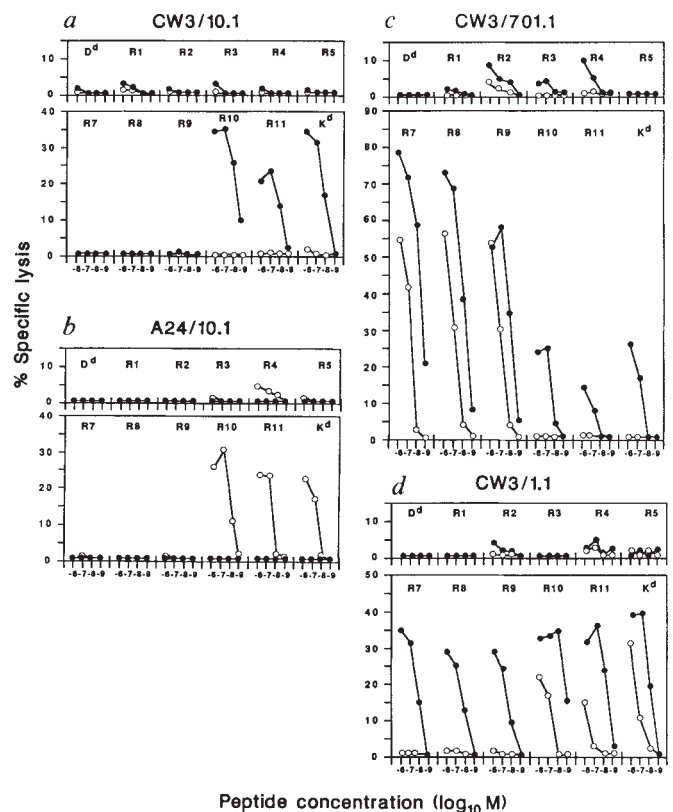


Fig. 1 Amino-acid sequence encoded by *K*^d/*D*^d recombinants and HLA peptides CW3 170–182 and A24 170–182.

express chimaeric *H-2* molecules composed of *K*^d sequences throughout the α 1 and α 2 domains and up to residues 184 and 191, respectively, were lysed comparably to control *K*^d transfectant target cells. Conversely, none of the clones recognized the HLA peptides on R1, R2, R3, R4 and R5 targets. Recognition of R7, R8 and R9 varied, depending on the combination of CTL clone and peptide tested. CTL clones CW3/10.1 (Fig. 2a) and A24/10.1 (Fig. 2b) that exclusively recognize CW3 or A24 peptides, respectively, on P815 cells and on *L-K*^d control targets failed to recognize these peptides on R7, R8 and R9 targets. In contrast, R7, R8 and R9 functioned as restriction elements for two other CTL clones, CW3/701.1 (Fig. 2c) and CW3/1.1 (Fig. 2d). The pattern of peptide recognition on these recombinants, however, was different from that found on the *K*^d transfectant controls. For example, CTL-CW3/701.1 normally recognizes only the CW3 peptide on *L-K*^d target cells, but clearly also recognized the A24 peptide on the R7, R8 and R9 recombinants, and only the CW3 peptide on R10 and R11. Moreover, clone CW3/1.1 recognizes both peptides on *L-K*^d target cells and on R10 and R11, but only the CW3 peptide on targets that expressed R7, R8 and R9 chimaeric molecules. Our results establish that R7 to R11 are all active in antigen presentation, but we cannot be certain that recombinants R1 to R5 are functional; lack of recognition of R1 to R5 could be due to a loss of either specificity or function.

Our data so far define two regions within the α 2 domain that each contain at least one residue critical for the function or specificity of the *K*^d molecule as a restriction element. These regions lie between the recombination end points of R5 and R7, and R9 and R10, as defined by shifts in the patterns of peptide recognition by four different CTL clones (Fig. 2). The region defined by R5 and R7 includes six *K*^d/*D*^d polymorphic residues (Fig. 1). Because the functionality of R5 is not established, we do not know whether several of these residues contribute to the function of the restriction element, or its specificity, or both.

First domain

H-2K^d: GPHSLRYFVTA VRPGLGEPRFIAVG YVDDTQFVR FSDADNPRFEPRAPWMEQEGPEYWE EQTORAKSDEQWFRVSLRTAQRYYNQSKG
 H-2D^d: -S-----F-----YME-----N-E-----E---Y---R-I-----RE-R---GN-S---D---L-----A-

Second domain

H-2K^d: GSHTFQRMFGCDVGS DWRLLRGYHQFAYDGRDYIALNEDLKTWTAADTAALITRRKWEQAGDAEY YRAYLEGECEVWLRRLRYELGNETLLRT
 H-2D^d: -L-W-A---E-G---W---C---M-Q---A-RD-----KN-A-----
 Peptide CW3 170-182 : ---KN-K---Q-A
 Peptide A24 170-182 : ---EN-K---Q-A

Third domain

H-2K^d: DSPKAHVITYHPRS QVDVTLRCWALGFYPADITLTWQLNGEDLTQDMELVETRPAGDGTGFKWAAVVPLGKEQNYTCHVHHKGLPEPLTLRWKL
 H-2D^d: -P-----H-R-PEG-----E---E-----S-----K---E-E-----GKE

Transmembrane and cytoplasmic domains

H-2K^d: PPSTVSTNTVIAVLVVLGAATVGTGA VFAVFMKMRRTGGKGVNYALAPGSQTS DLSLPDGKVMVHDPHSLA
 H-2D^d: EP-S--KT-----P-----VVIL--M-----R-----GD-----S--M-----C--

Fig. 2 Lysis of L cell K^d/D^d recombinant targets by K^d-restricted anti-HLA CTL clones in the presence of synthetic HLA peptides. Synthetic peptides corresponding to HLA-CW3 (●) or HLA-A24 (○) residues 170-182 (Fig. 1) were prepared and purified as described². Peptides were diluted in Dulbecco's modified Eagle's medium containing 5% fetal bovine serum and 100 µl volumes were added to wells of V-bottom microtitre plates. L cells were transfected with the HSV-*tk* gene²¹ and either a genomic clone of *H-2D^d* (ref. 22), a complementary DNA clone for *H-2K^d* (ref. 23), or the K^d/D^d recombinant clones shown in Fig. 1. Target cells (10⁶) were labelled with 150 µCi sodium [⁵¹Cr] chromate for 1 h at 37 °C and washed 3 times. Labelled target cells (2 × 10⁵ in 50 µl) were added to wells containing peptides. 2 × 10⁴ Cells from CTL clones a, CW3/10.1; b, A24/10.1; c, CW3/701.1; or d, CW3/1.1 in 50 µl were added to give an effector to target ratio of 10:1. After 6 h incubation at 37 °C, the plates were centrifuged and 100 µl supernatant was removed for counting. The % specific lysis was calculated as (experimental - spontaneous release)/(total - spontaneous release) × 100. There was no lysis in the absence of peptide.

The second critical region includes seven K^d/D^d polymorphic residues at positions 152, 155, 156, 173, 174, 177 and 184 (Fig. 1). In the latter case, the data show that at least one of these residues contribute to K^d specificity. It is interesting to note that the region 142-182 apparently contains determinants recognized by *H-2K^d*-specific alloreactive CTL (ref. 8). In addition, residues 155 and 156 contribute to epitope(s) of allogeneic CTL specific for a mutant of *H-2K^b* (ref. 9) and residue 152 of HLA-A3 is critical for influenza-specific HLA-A3-restricted CTL and for alloreactive CTL (ref. 10). In this context, it is remarkable that the presence of D^d residues at position 152, 155 and 156 still allows recognition of the CW3 peptide by CW3/701.1 and CW3/1.1 K^d-restricted CTL, but it must be emphasized that the specificity of recognition is altered for the A24 peptide.

The crystallographic structure of HLA-A2 has recently been solved^{11,12}. It has a peptide-binding pocket bordered by two α-helices laid on a platform made of β-strands. Given the very high amino-acid sequence homology of K^d, D^d and HLA-A2, it is tempting to assume that K^d and D^d have a similar three-dimensional structure. If so, the two critical regions identified here in K^d would map within the peptide pocket, one (defined by R5 and R7) at the bottom of the groove, and the other (defined by R9 and R10) in the longer of the two α-helices. Both could thus influence binding of the peptide as well as the shape of the bound peptide.

A key finding is that presentation of bound peptides by the chimaeric restriction elements is apparently not sufficient for their recognition by certain CTL clones. For example, CW3 peptide is most probably bound by R7, R8, R9, R10 and R11, as CTL clones CW3/701.1 and CW3/1.1 lyse all of those recombinants together with the peptide. Nevertheless, clone CW3/10.1 does not lyse R7, R8 and R9 although R10 and R11 are lysed.

Furthermore, the patterns of cross-reactivity of CW3 and A24 peptides with respect to the CTL clones depend on which chimaeras are used. For clone CW3/1.1, recognition of the A24 peptide, but not that of CW3, is lost on R7, R8 and R9. Remarkably, clone CW3/701.1, that exclusively recognized the CW3 peptide on *L-K^d* target cells, recognized both CW3 and A24 peptides on R7, R8 and R9 recombinants. In contrast, this clone could not lyse R10 and R11 with added A24 peptide, although the latter presumably binds to R10 and R11 chimaeric molecules, as indicated by the reactivities of clones A24/10.1 and CW3/1.1.

Such results are reminiscent of data obtained with class II MHC molecules, for example those of Heber-Katz *et al.*¹³, who analysed the response of class II-restricted T-cell hybridomas to homologous cytochrome *c* fragments from pigeon or moth on antigen-presenting cells from different MHC haplotypes. When tested on antigen-presenting cells that expressed hybrid MHC molecules made of chains originating from distinct MHC haplotypes, the pattern of cytochrome *c* fragment recognition was found to be altered. These and other results¹⁴, together with our data, suggest that for both class I- and class II-restricted T cells, the epitopes recognized may be in part determined by the presenting MHC molecule. The most likely interpretations are either that the T-cell antigen receptor (TcR) binding site interacts with residues from both the antigenic peptide and the MHC restriction element, or that the latter participates in the conformation of the presented peptide¹⁴, or both.

The molecular details of the presumed interaction between MHC molecules, peptides, and antigen-specific TcRs are not yet understood. The binding of several known immunogenic peptides to purified class II molecules has been demonstrated¹⁵⁻¹⁷. A high degree of correlation was found between the apparent binding constants of different peptides for specific

class II molecules and their known restriction patterns^{15,16,18}. Moreover, peptides recognized in the context of the same restriction element could effectively compete with each other for *in vitro* binding and for functional recognition¹⁸⁻²⁰, suggesting that MHC class II molecules contain a single antigen-binding site. Similar experiments have not yet been reported for MHC class I molecules, but the demonstration that antigens recognized by class I restricted CTL can be mimicked by synthetic peptides^{1,2}, together with the presumptive structural homologies between class I and class II molecules^{11,12}, suggest that the general features involved in antigen recognition may be similar for both classes of MHC molecules. Further combinations of peptides, chimaeric class I MHC molecules, and CTL clones should be useful in resolving this issue.

We would like to thank Drs J.-C. Cerottini, G. Corradin and P. Pala for discussion, K. Muhlethaler for technical assistance, and P. Brunet for help in the preparation of the manuscript.

Received 24 August; accepted 29 October 1987.

1. Townsend, A. R. M. *et al.* *Cell* **44**, 959-968 (1986).
2. Maryanski, J. L., Pala, P., Corradin, G., Jordan, B. R. & Cerottini, J.-C. *Nature* **324**, 578-579 (1986).
3. Abastado, J. P., Darche, S., Godeau, F., Cami, B. & Kourilsky, P. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6496-6500 (1987).
4. Maryanski, J. L., Accolla, R. S. & Jordan, B. J. *Immun.* **136**, 4340-4347 (1986).
5. Burgert, H.-G., Maryanski, J. L. & Kvist, S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1356-1360 (1987).
6. Maryanski, J. L. *et al.* *Proc. 18th International Leukocyte Culture Conf.* (in the press).
7. Abastado, J. P. *et al.* *J. exp. Med.* **166**, 327-340 (1987).
8. Scholler, J., Shimonkevitz, R., MacDonald, H. R. & Kvist, S. *J. exp. Med.* **164**, 1823-1834 (1986).
9. Sherman, L. A. *Nature* **297**, 511-513 (1982).
10. Cowan, E. P., Jordan, B. R. & Coligan, J. E. *J. Immun.* **135**, 2835-2841 (1985).
11. Bjorkman, P. J. *et al.* *Nature* **329**, 506-512 (1987).
12. Bjorkman, P. J. *et al.* *Nature* **329**, 512-518 (1987).
13. Heber-Katz, E. *et al.* *J. exp. Med.* **155**, 1086-1099 (1982).
14. Claverie, J. M. & Kourilsky, P. *Ann. Immun. (Inst. Pasteur)* **137D**, 425-442 (1986).
15. Babbitt, B. P., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. R. *Nature* **317**, 359-361 (1985).
16. Buus, S. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 3968-3971 (1986).
17. Buus, S., Sette, A., Colon, S. M., Jenis, D. M. & Grey, H. M. *Cell* **47**, 1071-1077 (1986).
18. Buus, S., Sette, A., Colon, S. M., Miles, C. & Grey, H. M. *Science* **235**, 1353-1358 (1987).
19. Guillet, J.-G., Lai, M. Z., Briner, T. J., Smith, J. A. & Geftner, M. L. *Nature* **324**, 260-262 (1986).
20. Guillet, J. G. *et al.* *Science* **235**, 865-870 (1987).
21. Colbère-Garapin, F., Chousterman, S., Horodniceanu, F., Kourilsky, P. & Garapin, A. C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3755-3759 (1979).
22. Greenstein, J. L., Malissen, B. & Burakoff, S. J. *J. exp. Med.* **162**, 369-374 (1985).
23. Cochet, M. *et al.* *Immunogenetics* **24**, 267-274 (1986).
24. Pala, P. *et al.* *J. Immun.* (in the press).
25. Maryanski, J. L., Pala, P., Cerottini, J.-C. & Corradin, G. *J. exp. Med.* (in the press).

Anti-cachectin/TNF monoclonal antibodies prevent septic shock during lethal bacteraemia

Kevin J. Tracey*‡, Yuman Fong*, David G. Hesse*, Kirk R. Manogue‡, Annette T. Lee‡, George C. Kuo†, Stephen F. Lowry*§ & Anthony Cerami‡

* Laboratory of Surgical Metabolism, Department of Surgery, New York Hospital-Cornell University Medical Center, New York, New York 10021, USA

† Chiron Research Laboratories, Emeryville, California 94608, USA

‡ Laboratory of Medical Biochemistry, Rockefeller University, New York, New York 10021, USA

Bacterial infection of the mammalian bloodstream can lead to overwhelming sepsis, a potentially fatal syndrome of irreversible cardiovascular collapse (shock) and critical organ failure. Cachectin, also known as tumour necrosis factor, is a macrophage-derived peptide hormone released in response to bacterial lipopolysaccharide, and it has been implicated as a principal mediator of endotoxic shock, although its function in bacterial sepsis is not known. Anaesthetized baboons were passively immunized against

Table 1 Survival after pre-treatment with monoclonal anti-cachectin F(ab')₂ fragments during live gram-negative bacteraemia

Group	n	Survivors	Mean survival time (h ± s.e.m.)
Controls	6	0	13.3 ± 3.5
Antibody (-1 h)	3	0	17.4 ± 2.6
Antibody (-2 h)	3	3	sacrificed*

Female *Papio anubis* baboons (13.1 ± 0.6 kg) from the Charles River Primate Center (Port Jefferson, New York) were housed for a minimum of two weeks in the animal care facilities of the New York Hospital-Cornell University Medical Center (CUMC). The experimental protocol was reviewed and approved by the Institutional Animal Care and Use Committee at CUMC. Animals were free of infections or parasites, with haematocrits exceeding 36%. Food intake was restricted for 12 h before the experiment. Each baboon was immobilized with ketamine (14 mg kg⁻¹, i.m.) and the cephalic vein cannulated with a 20-gauge Teflon catheter; sodium pentobarbital was administered (2.0 mg kg⁻¹ intravenously, at 30 min intervals) to maintain anaesthesia. The trachea was intubated with a cuffed tube and animals maintained spontaneous respirations on warming blankets. The superficial femoral vessels were exposed aseptically in one hindlimb; the artery was cannulated and connected to a pressure transducer used for the continuous monitoring of blood pressure and heart rate; the vein was cannulated with a flow-directed, thermolabile catheter (American Edwards Laboratories) which was guided into the pulmonary artery for measurement of pulmonary artery pressure and cardiac output. Control baboons were injected with carrier vehicle (5 ml 0.9% saline per kg, intravenous) by constant infusion over 30 min at *t* = -1 h (*n* = 3) or *t* = -2 h (*n* = 3). Experimental animals received anti-cachectin antibody F(ab')₂ fragments (10 mg F(ab')₂/kg in 5 ml 0.9% saline per kg) administered by constant intravenous infusion for 30 min beginning at *t* = -1 h or *t* = -2 h as indicated. Investigators were not informed whether infused carrier contained F(ab')₂. Lyophilized *E. coli* 086:B7 (from G. T. Shires) were used to inoculate slant cultures on tryptic soy broth agar²³; viability counts of the inoculum were determined by standard dilution techniques. At time zero, baboons received an infusion of 1.2 ± 0.3 × 10¹¹ live bacteria per kg body weight delivered into the aorta by constant infusion for 30 min; animals were maintained under anaesthesia and monitored continuously for 10 h, vital signs being recorded every 15 min, pulmonary artery pressures every 30 min, and cardiac output hourly.

* Animals surviving longer than 48 h were considered survivors and were subsequently sacrificed for necropsy at 2 d, 4 d, or 7 d.

endogenous cachectin and subsequently infused with an LD₁₀₀ dose of live *Escherichia coli*. Control animals (not immunized against cachectin) developed hypotension followed by lethal renal and pulmonary failure. Neutralizing monoclonal anti-cachectin antibody fragments (F(ab')₂) administered to baboons only one hour before bacterial challenge protected against shock, but did not prevent critical organ failure. Complete protection against shock, vital organ dysfunction, persistent stress hormone release and death was conferred by administration of antibodies 2 h before bacterial infusion. These results indicate that cachectin is a mediator of fatal bacteraemic shock, and suggest that antibodies against cachectin offer a potential therapy of life-threatening infection.

Sepsis can occur as a primary disease process or secondarily complicate other disease states, and is associated with a high mortality, despite antibiotic therapy^{1,2}. Immunocompromised patients, such as those with malignancy, trauma or AIDS (acquired immune deficiency syndrome), are particularly susceptible. The toxicity of gram-negative bacteraemia was originally attributed to endotoxin/lipopolysaccharide (LPS), because many of the deleterious consequences of bacterial sepsis can be induced by LPS administration³, but endogenous cytokines, rather than LPS itself, are now known to mediate the homeostatic abnormalities during endotoxaemia⁴⁻⁷. Serum levels of cachectin achieve nanomolar concentrations in experimental models of lethal endotoxaemia^{7,8}: when infused into healthy animals, recombinant human-cachectin can induce haemodynamic collapse, counter-regulatory hormone release,

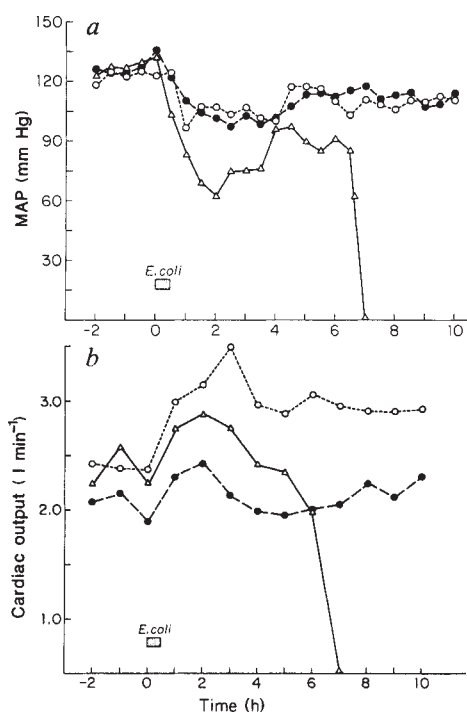


Fig. 1 *a*, Mean arterial blood pressure (MAP) and *b*, cardiac output in individual baboons during experimental primate bacteraemia. Intravenous *E. coli* was infused during the time indicated by the solid bars. The control (open triangle) received saline at $t = -2$ h; anti-cachectin $F(ab')_2$ was given at either $t = -1$ h (open circles) or $t = -2$ h (filled circles), as indicated. To more closely approximate the typical clinical resuscitation of septic human patients, and to minimize intravascular dehydration (which can confuse the interpretation of cardiovascular responses), Ringer's lactate solution was administered intravenously at a rate of 2 ml per kg-h throughout the study period. Additional Ringer's lactate was administered as a bolus (10 ml kg⁻¹ over 15 min) if the pulmonary capillary wedge pressure (PCWP) fell below 2 mm Hg. PCWP was measured immediately after completion of the fluid bolus and the infusion repeated if necessary. Intravenous fluid administration was suspended if the PCWP was maintained at ≥ 3 mm Hg.

and critical organ injury in a pattern which is strikingly similar to that observed during lethal endotoxaemia^{9,10}. Polyclonal antibodies raised against cachectin prevented death during subsequent endotoxaemia in mice, suggesting that cachectin might mediate lethal endotoxaemia¹¹. Although LPS initiates the pathology of gram-negative sepsis, and cachectin can reproduce lethal endotoxaemia, it remains unclear whether cachectin is a necessary stimulus to septic shock.

In the present study, we produced lethal bacteraemia in baboons and specifically blocked endogenous cachectin by passive immunization with monoclonal anti-cachectin $F(ab')_2$ fragments. Anaesthetized baboons were monitored for blood pressure, pulmonary arterial pressure and cardiac output as described in the legend to Table 1, with one control and one $F(ab')_2$ antibody-treated animal studied simultaneously. Solutions were administered intravenously either 1 or 2 h before an LD₁₀₀ dose of live *E. coli*^{12,13}, given by intra-aortic infusion. All animals surviving the 10-h study period were treated with an antibiotic (5 mg gentamicin kg⁻¹ by intramuscular injection twice daily).

Maximal plasma cachectin levels (21.5 ± 3.8 ng ml⁻¹, mean $\pm$ s.e.m.) in the controls were observed between 1.5–2.5 h following the *E. coli* infusion, as assayed by a cachectin enzyme-linked immunosorbent assay (ELISA) (ref. 14). The concentration of serum cachectin returned from nanomolar to below detectable levels (34 pg ml⁻¹) 4–6 h after bacterial infusion; this rapid disappearance is consistent with previous observations during experimental endotoxaemia^{8,14}, and the magnitude of the response is sufficient to induce lethal shock¹⁰. In contrast, the administration of neutralizing anti-cachectin antibodies attenuated the circulating cachectin response (Table 2). The non-immunized (control) baboons rapidly succumbed to the lethal effects of gram-negative bacteraemia (Table 1). The infusion of live bacteria in the unprotected controls typically induced an acute decrease in both blood pressure and in pulmonary capillary wedge pressure, and an increase in heart rate (from 120 beats per min at baseline to >190 beats per min at 2 h) (Fig. 1). Cardiac output increased transiently with administration of intravenous fluid and in response to endogenous adrenaline release (Table 2). With the development of sepsis in controls, there was an inexorable decline in cardiac output, worsening shock, vital organ injury, and anuric renal failure associated

Table 2 Systemic hormone and leukocyte responses to live *E. coli* bacteraemia after pre-treatment with anti-cachectin $F(ab')_2$

	<i>n</i>	Sample time (h)*	Cachectin (ng ml ⁻¹)	Adrenaline (pg ml ⁻¹)	Noradrenaline (pg ml ⁻¹)	Glucagon (pg ml ⁻¹)	Leukocytes ($\times 10^3$ cells μ l ⁻¹)
Controls	6	-2	<0.04	165 $\pm$ 27	382 $\pm$ 40	115 $\pm$ 17	13.2 $\pm$ 0.8
		2	16.7 $\pm$ 0.3	973 $\pm$ 302	2,344 $\pm$ 960	289 $\pm$ 34	1.1 $\pm$ 0.1
		8	<0.04	3,012 $\pm$ 861	3,411 $\pm$ 1,156	>1200	1.2 $\pm$ 0.1
Antibody (-1 h)	3	-2	<0.04	129 $\pm$ 26	958 $\pm$ 599	134 $\pm$ 48	9.3 $\pm$ 0.9
		2	1.2 $\pm$ 0.9	1,785 $\pm$ 1,135	2,591 $\pm$ 934	293 $\pm$ 92	1.0 $\pm$ 0.2
		8	<0.04	3,202 $\pm$ 1,087	3,934 $\pm$ 676	>1200	1.4 $\pm$ 0.2
Antibody (-2 h)	3	-2	<0.04	191 $\pm$ 25	419 $\pm$ 87	156 $\pm$ 16	12.1 $\pm$ 1.1
		2	<0.04	550 $\pm$ 162	491 $\pm$ 322	574 $\pm$ 14	3.6 $\pm$ 1.2
		8	<0.04	160 $\pm$ 16	372 $\pm$ 110	365 $\pm$ 123	8.2 $\pm$ 4.0

Murine monoclonal antibodies against recombinant human cachectin were produced in mouse ascites fluid and immunoglobulin (IgG) prepared by ammonium sulphate fractionation. The $F(ab')_2$ fragments were generated by pepsin digestion of IgG (10 mg ml⁻¹) with pepsin (0.02 mg ml⁻¹) in citrate buffer (pH 3.5) for 3 h at 37 °C (ref. 24). $F(ab')_2$ was purified by DEAE-Sepharose column chromatography and dialysed exhaustively against phosphate-buffered saline (pH 7.4). Neutralizing activity of the purified $F(ab')_2$ was determined in a standard bioassay based on L929-cell cytotoxicity^{25,26}. Purified $F(ab')_2$ completely neutralized at least 50 ng ml⁻¹ of recombinant human cachectin (10^8 U mg⁻¹) at an antibody concentration of 10 μ g ml⁻¹. The purified $F(ab')_2$ (10 μ g ml⁻¹) also completely neutralized L929 cell cytotoxicity of baboon plasma collected 2.5 h after bacteraemia during a maximal cachectin response (L929 cell cytotoxicity equivalent to 100 ng ml⁻¹ r-human cachectin). The endotoxin/LPS content of the purified $F(ab')_2$ solution was less than 0.25 ng LPS per mg protein, as assayed by the Limulus amoebocyte lysate test. Baseline arterial blood was collected at $t = -2$ h after the pulmonary capillary wedge pressure had stabilized at 2–3 mm Hg. Samples were placed in ice, centrifuged (2,100 r.p.m., 4 °C, 20 min) and stored at -70 °C. Heparinized plasma was assayed for cachectin by ELISA (ref. 14); adrenaline and noradrenaline were assayed by radioenzymatic assay²⁷; plasma for glucagon was collected with EDTA and aprotinin before analysis by radioimmunoassay²⁸. Data recorded as mean $\pm$ standard error.

* Samples were collected at the following times: baseline specimens (before the infusion of antibody or saline) at $t = -2$ h; after the start of the bacterial infusion ($t = 2$ h); and $t = 8$ h, or immediately before death.

with a twofold increase in serum creatinine, followed by death from pulmonary oedema. These pathophysiological events are characteristic of lethal, hypodynamic sepsis^{12,13}.

Passive immunization of baboons with anti-cachectin antibody infused 1 h before bacterial challenge conferred beneficial cardiovascular effects (Fig. 1) but not complete protection against critical organ injury (Table 1). In contrast to the controls, blood pressure did not fall acutely after bacteraemia, but was maintained by a compensatory increase in heart rate (up to 175 beats per min) and cardiac output. Although acute cardiovascular collapse and shock were not observed, serious renal injury did occur, as evidenced by twofold increases of serum creatinine at 8 h and the development of anuria. Each of the baboons immunized 1 h before bacterial challenge developed fatal pulmonary oedema.

Should uniform distribution and tissue penetration of the antibody not have occurred within 1 h, we investigated protection using earlier passive immunization with anti-cachectin F(ab')₂. When antibody was administered 2 h before *E. coli*, normal blood pressure was maintained during bacteraemia, significant tachycardia occurred only transiently, and shock was prevented. Neither did vital organ dysfunction occur or renal failure develop, and there was no evidence of pulmonary oedema. Animals recovered from anaesthesia, were active and healthy, and resumed eating within 24 h. No evidence of persisting sepsis or tissue injury was observed by physical examination or routine complete blood count until the time that animals were killed for necropsy. Thus, early passive immunization with monoclonal anti-cachectin antibodies protected against the effects of bacterial sepsis.

These results with anti-cachectin antibodies were not due to increased bacterial clearance in the immunized animals, or to bacteriocidal properties of the antibody solution: the viability of circulating bacteria 2 h after *E. coli* infusion was similar in the control ($3.1 \pm 0.5 \times 10^4$ CFU per ml) and immunized baboons ($3.9 \pm 0.8 \times 10^4$ CFU per ml), and decreased comparably in both groups within 4 h ($1.3 \pm 0.6 \times 10^3$ CFU per ml). During this period, the inhibition of cachectin by F(ab')₂ resulted in improved cardiac output, suggesting that cachectin is necessary to provoke septic shock. The administration of F(ab')₂ alone without bacterial challenge did not significantly change blood pressure, cardiac output or hormone counter-regulatory release over the 10-h monitoring period. Recovery from anaesthesia was uneventful, normal activity and food intake was resumed overnight and post-mortem examination was normal. The improved cardiovascular function in the antibody-treated bacteraemic animals cannot be attributed to differences in hydration status, because similar amounts of fluid were administered to all animals in accordance with a predetermined protocol (data not shown).

During sepsis, the systemic release of catabolic stress hormones in part mediates the maintenance of cardiovascular tone and mobilization of host energy stores¹⁵. We observed persistent increases in circulating adrenaline, noradrenaline, and glucagon in all animals succumbing to bacterial challenge (Table 2), but early pre-treatment with antibodies blunted the magnitude of later (8 h) counter-regulatory hormone responses in all survivors. We have previously shown that the infusion of recombinant cachectin produces hypotension and diminished cardiac output despite endogenous catecholamine production (which would normally increase cardiac output)¹⁰. As antibody administration is associated with improved cardiac output, the present data also suggest that sepsis-associated myocardial depression is due in part to cachectin.

It appears then that the normal injurious host responses elicited by overwhelming bacteraemia can be prevented if the effects of cachectin are blocked. The death of the infected host could result from an overexpression of cachectin during a normally beneficial immune response, somewhat analogous to anaphylactic shock, during which protective immune responses

are capable of inducing shock and death. It is probable that the provocative toxicity of cachectin arises in part from a direct effect on normal tissues, and in part from the release of other humoral factors. Complement activation, release of the interleukins, interferons and platelet-activating factor, and the induction of other cytokines would be expected to amplify and broaden the range of host responses¹⁶⁻¹⁹.

Experimental models of endotoxaemia and sepsis perhaps differ from the indolent and progressive syndrome of multiple system organ failure in human patients, but cachectin has been implicated in human infections^{20,21}. Levels of cachectin are frequently elevated during acute meningococcal sepsis, when those patients with the highest cachectin levels die²². The prophylactic use of anti-cachectin antibodies in patients at high risk for overwhelming infection could protect against septic shock. Further experiments are needed to determine whether the administration of anti-cachectin antibodies during persisting infection or indolent sepsis will be beneficial.

The authors wish to thank E. Diane Spurgeon, Robin Solomon and He Wei for excellent technical assistance; and Harvey Cohen, Thomas Okarma and Patricia Olson for helpful discussions.

Received 27 August; accepted 19 October 1987.

1. Kreger, B. E. W., Craven, D. E. & McCabe, W. R. *Am. J. Med.* **68**, 344-355 (1980).
2. Ziegler, E. J. *et al. New Engl. J. Med.* **307**, 1225-1230 (1982).
3. Gilbert, R. P. *Physiol. Rev.* **40**, 245-279 (1960).
4. Beutler, B. & Cerami, A. *Nature* **320**, 584-588 (1986).
5. Beutler, B. & Cerami, A. *New Engl. J. Med.* **316**, 379-385 (1987).
6. Tracey, K. J., Lowry, S. F. & Cerami, A. in *TNF and Related Cytotoxins. CIBA Found. Symp.* **131**, 88-108 (Wiley, London, 1987).
7. Tracey, K. J., Lowry, S. F. & Cerami, A. *J. infect. Dis.* (in the press).
8. Beutler, B., Milsark, I. W. & Cerami, A. *J. Immun.* **135**, 3972-3977 (1985).
9. Tracey, K. J. *et al. Science* **234**, 470-474 (1986).
10. Tracey, K. J. *et al. Surg. gynecol. Obstet.* **164**, 415-422 (1987).
11. Beutler, B., Milsark, I. W. & Cerami, A. *Science* **229**, 869-871 (1985).
12. Cryer, P. E., Herman, C. M. & Sode, J. *Ann. Surg.* **174**, 91-100 (1971).
13. Coalson, J. J., Hinshaw, L. B., Guenter, C. A., Berrel, E. L. & Greenfield, L. J. *Lab. Invest.* **32**, 561-569 (1975).
14. Hesse, D. G. *et al. Surg. gynecol. Obstet.* (in the press).
15. Lowry, S. F. in *Advances in Host Defense Mechanisms* Vol. 6 (eds Gallin, J. I. & Fauci, A. S.) 169-190 (Raven, New York, 1986).
16. McCabe, W. R. *New Engl. J. Med.* **288**, 21-23 (1973).
17. Dinarello, C. A. *et al. J. exp. Med.* **163**, 1433-1450 (1986).
18. Darius, H., Lefer, D. J., Smith, J. B. & Lefer, A. M. *Science* **232**, 58-60 (1986).
19. Chang, S. W., Feddersen, C. O., Henson, P. M. & Voelkel, N. F. *J. clin. Invest.* **79**, 1498-1509 (1987).
20. Scuderi, P. *et al. Lancet* **ii**, 1364-1365 (1986).
21. Tracey, K. J. *et al. Surg. Forum* **37**, 13-15 (1986).
22. Waage, A., Halstensen, A. & Espevik, T. *Lancet* **i**, 355-357 (1987).
23. Pool, J. L. *Surg. gynecol. Obstet.* **132**, 1-9 (1971).
24. Parham, P. J. *Immun.* **131**, 2895-2899 (1983).
25. Peters, P. M. *et al. J. Immun.* **137**, 2592-2598 (1986).
26. Carswell, E. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **72**, 3666-3670 (1975).
27. Passon, P. G. & Peuler, J. D. *Analyt. Biochem.* **51**, 618-631 (1973).
28. Aquilar-Parada, A., Eisentraut, E. M. & Unger, R. H. *Diabetes* **18**, 717-723 (1969).

γ Interferon, CD8⁺ T cells and antibodies required for immunity to malaria sporozoites

Louis Schofield*, Jaime Villalquira*, Arturo Ferreira†, Huub Schellekens‡, Ruth Nussenzweig* & Victor Nussenzweig†

Departments of Medical and Molecular Parasitology*, and Pathology†, New York University School of Medicine, New York, New York 10016, USA

‡ Primate Center TNO, Rijswijk, The Netherlands

This study was designed to test the hypothesis that T-cell effector mechanisms are required for protective immunity to malaria sporozoites. Administration of neutralizing monoclonal antibodies against γ interferon (γ IFN) to immune hosts, reversed sterile immunity to sporozoite challenge, by allowing the growth of

exoerythrocytic forms (EEF) and thus the development of parasitaemia. Immune animals also developed infections when depleted *in vivo* of their suppressor/cytotoxic T cells expressing the CD8 antigen (CD8⁺) but not when depleted of helper T cells expressing CD4 antigen (CD4⁺), before sporozoite challenge. Passive transfer of immune immunoglobulin alone, or adoptive transfer of immune T cells alone, conferred partial protection to naive recipients. Transfer of both immune components resulted in significantly greater protection. This transferred immunity was reversed by the *in vivo* neutralization of γ IFN. Thus, sterile immunity to sporozoite challenge requires the neutralization of sporozoites by antibodies and the inhibition of EEF development by γ IFN with the participation of CD8⁺ cells.

Vaccination of hosts with attenuated malaria sporozoites confers protective immunity to viable sporozoite challenge¹. This immunity is mediated in part by antibodies to the circumsporozoite (CS) protein²⁻⁴. As immunity can be induced in B-cell depleted mice⁵, however, T cells may also contribute to protection. It is difficult, however, to envisage cellular mechanisms which could neutralize a large sporozoite inoculum in the short time before hepatocyte invasion.

Recent studies show that a novel T-cell mediated immune mechanism may act not against sporozoites, but against the developing liver-stage parasites: recombinant γ IFN at very low concentrations inhibits the development of EEF within hepatocytes, *in vivo*^{6,7} and *in vitro*⁸. In this study, we tested the hypothesis that γ IFN, released from sensitized T cells during sporozoite challenge, inhibits the development of EEF, and is required for sterile immunity. We also investigated whether helper or cytotoxic T cells are required for immunity to sporozoite challenge.

Immunized rats were challenged with *Plasmodium berghei* sporozoites. Following challenge, the animals received either monoclonal antibody (mAb) DB-1, which neutralizes rat and mouse γ IFN⁹, or an irrelevant control mAb. After 44 h the level of EEF development was determined by DNA hybridization¹⁰. As shown in Fig. 1, naive animals receiving DB-1 or the control mAb showed similar high levels of EEF growth. As expected, immunized animals receiving the irrelevant mAb showed no detectable levels of EEF DNA. Immunized animals receiving mAb DB-1, however, showed substantial levels of EEF development, that is, 43.5% of control levels with inocula of 2.5×10^4 or 5×10^4 sporozoites. A similar proportion of EEF was rescued from destruction by endogenous γ IFN activity in immune hosts when DB-1 was administered 2.5 h after challenge. As sporozoites can invade hepatocytes within 2 min of inoculation¹¹, and are cleared from the circulation within 30 min¹², the targets of endogenous γ IFN are indeed the developing intrahepatocytic EEF.

We extended our experiments to determine whether *in vivo* neutralization of γ IFN, or *in vivo* depletion of immune T cells, would reverse sterile immunity and allow the development of blood-stage parasites. Mice that had been immunized according to established protocols remained protected when challenged with 5×10^3 *P. berghei* sporozoites, followed by 1 mg of irrelevant control mAb. Blood-stage infection, however, developed in immunized animals that received 1 mg of mAb DB-1 either immediately or 2.5 h after challenge (Table 1, Expt 1). In another experiment, mice which had been hyperimmunized with repeated injections of large numbers of X-irradiated sporozoites resisted challenge doses of 2×10^4 – 1.5×10^5 sporozoites. In parallel experimental groups receiving DB-1 after challenge with 1.5×10^5 , 10^5 or 5×10^4 sporozoites, most animals developed blood-stage parasitaemia. The anti-sporozoite-antibody titres in the hyperimmunized groups were high (24,576) which may account for their resistance to the lower challenge dose of 2×10^4 sporozoites (Table 1, Expt 2).

γ IFN can be produced by helper and suppressor/cytotoxic T cells. To determine if either T-cell subset is involved in mediating protection, groups of immunized mice received MABs

Table 1 Effect of γ IFN neutralization on immunity to sporozoite challenge

Immune status	Challenge dose	No. infected/total		Day of patency (range)		Reciprocal IFA titre
		DB-1	Control mAb	DB-1	Control mAb	
Expt 1						
Naive	5,000*	5/5	5/5	4.0 (4.0)	4.0 (4.0)	32
Immune	5,000*	5/5	0/5	5.6 (4–7)	—	6,144
Immune	5,000†	4/5	0/5	6.0 (5–7)	—	6,144
Expt 2						
Naive	20,000	5/5	5/5	4.0 (4.0)	4.0 (4.0)	32
Hyperimmune	20,000	0/5	0/5	—	—	24,576
Hyperimmune	50,000	2/5	0/5	6.5 (6–7)	—	24,576
Hyperimmune	100,000	4/5	0/5	6.75 (6–7)	—	24,576
Hyperimmune	150,000	4/5	0/5	5.5 (5–6)	—	24,576

Neutralization of endogenous γ IFN reverses immunity to sporozoite challenge. Antibody titres of pooled sera determined by IFA using glutaraldehyde-fixed sporozoites. Mice were challenged with variable doses of sporozoites. Test groups received either control mAb or mAb DB-1. The appearance of parasitaemia (patency) was determined by daily examination of thin blood films. Experiment 1; test groups immunized with 6×10^4 sporozoites followed by two boosts of 10^4 sporozoites at two-week intervals; * received mAbs immediately after challenge; † received mAbs 2½ h after challenge. Experiment 2; test groups hyperimmunized with 10^5 sporozoites followed by four boosts of 10^5 , 5×10^4 , 5×10^4 and 5×10^4 sporozoites at three-week intervals. Naive controls and test groups received mAbs immediately after challenge. Comparison of control and treatment groups in both experiments showed a significant level of heterogeneity ($P < 0.01$), by Kruskal-Wallis test. The level of significance was adjusted using the Bon-Ferroni transformation.

with lytic activity for CD4⁺ or CD8⁺ T cells¹³. As shown in Table 2, the elimination of CD8⁺ T cells in fully immune animals rendered them susceptible to challenge with 5×10^3 sporozoites. In contrast, destruction of CD4⁺ T cells did not reverse host immunity.

Passive and adoptive transfer experiments were undertaken to clarify the relative contributions of humoral and cellular mechanisms to protective immunity. Naive mice received purified IgG or splenic T lymphocytes from immune or naive donors, or a combination of immune IgG and T cells. In contrast to previous studies^{14,15}, the mice did not receive booster immunizations with irradiated sporozoites after transfer of spleen cells. As shown in Table 3, control mice which received either nonimmune IgG or naive T lymphocytes soon developed parasitaemia. Those receiving immune IgG alone, or immune T lymphocytes alone, were partially protected against infection, that is, the prepatent period was significantly prolonged, indicating a substantial reduction in the infectivity of sporozoites or in the development of EEF. Mice receiving both immune IgG and immune T lymphocytes showed the greatest degree of protection (60%). This transferred immunity was reversed by administration of mAb DB-1 to recipients ($P < 0.001$). Clearly, challenge with a small number of sporozoites (10^3), similar to

Table 2 Effect of CD8⁺ depletion on immunity to sporozoite challenge

Immune status	No. infected/total	Mean day of patency (range)
Immunized, untreated	0/10	—
Immunized, CD4 ⁺ depleted	0/5	—
Immunized, γ IFN depleted	4/5	5.25 (5–6)
Immunized, CD8 ⁺ depleted	5/5	4.4 (4–5)
Naive	5/5	4.8 (4–5)

Depletion of CD8⁺, but not CD4⁺, T cells reverses immunity to sporozoite challenge. Immune animals were inoculated intravenously with mAbs with lytic activity against CD8⁺ or CD4⁺ T cells, or DB-1, and were challenged with 5×10^3 *P. berghei* sporozoites. The degree of T-cell subset depletion *in vivo* was determined by surface immunofluorescence assay on splenic lymphocytes from parallel T-cell depleted and control animals (92% and 94% depletion for CD8⁺ and CD4⁺ subsets, respectively).

Table 3 Passive and adoptive transfer of immunity to sporozoites and EEF

Treatment	No. infected/total	Mean of patency (range)	Reciprocal IFA titre
Expt 1			
Non-immune T cells	5/5	4.2 (4-5)	16
Immune T cells	5/5	5.8 (5-6)	32
Immune T cells + mAb DB-1	5/5	4.2 (4-5)	32
Expt 2			
Nonimmune IgG	4/5	4.0 (4)	64
Non-immune T cells	5/5	4.2 (4-5)	32
Immune IgG	4/5	5.5 (5-6)	2,048
Immune T cells	4/5	5.75 (4-7)	128
Immune Ig + immune T cells	2/5	7.5 (7-8)	3,072
Immune Ig + immune T cells + mAb DB-1	5/5	5.0 (5)	3,072

Donor A/J mice were immunized with X-irradiated sporozoites. Affinity-purified serum IgG and splenic T lymphocytes (2×10^7) purified on Lymphocyte M-separation medium (Accurate), followed by panning on affinity-purified goat anti-mouse IgG-coated plates, were inoculated intravenously into naive recipient A/J mice 36 h before challenge with 10^3 (Expt 1) or 3.5×10^3 (Expt 2) viable sporozoites. Some groups received mAb DB-1, or control mAb, immediately following challenge. Patency and antibody titres were determined as described (Table 1). Comparison of control and treatment groups was by Kruskal-Wallis test ($P < 0.01$). $P < 0.001$ for comparison of the differences in patency between recipients of both immune antibodies and immune T cells, and recipients of immune antibodies, immune T cells and mAb DB-1, by Mann-Whitney (Wilcoxon) test.

that inoculated by the bite of infected mosquitoes, is sufficient to induce a γ IFN response (Table 3, expt 1).

CD8⁺ T cells may exert their effect in two ways. They may respond to the presentation of parasite antigens in the context of Class I major histocompatibility complex (MHC) glycoproteins with the production of γ IFN which then acts in a hormonal fashion against EEF. Binding of the lymphokine to surface receptors on infected hepatocytes destroys EEF *in vitro* without the participation of immune effector cells⁸. In support of this interpretation, the protection afforded by adoptive transfer of splenic T cells was reversed by mAb DB-1 (Table 3). In addition CD8⁺ T cells may be directly cytotoxic for infected hepatocytes which present processed antigens of sporozoite or EEF origin.

Although some sporozoites can evade anti-sporozoite antibodies and become established within the host liver (Fig. 1), γ IFN alone may not be sufficient to provide sterile immunity, as very high levels are required to eradicate EEF⁶⁻⁸. Thus, whereas each component may be insufficient to mediate complete protection, they can act additively (Table 3). In another study, rodents were immunized with a synthetic peptide corresponding to the repeat domain of the CS protein of *P. berghei*. High levels of antibodies to sporozoites were induced without detectable peptide-specific cellular responses. Extensive protection was observed upon challenge with a small number of sporozoites¹⁶. Thus the relative contribution of humoral and cellular mechanisms to sporozoite/EEF immunity depends upon the interaction of at least three variables, namely the titre of neutralizing antibodies, the degree and type of cellular response to parasite antigen and the challenge dose of viable sporozoites itself, and will therefore vary with choice of antigen and method of vaccination.

As *in vivo* destruction of CD4⁺ T cells did not affect host immunity, and CD4⁺-intact, CD8⁺-deleted animals were suscep-

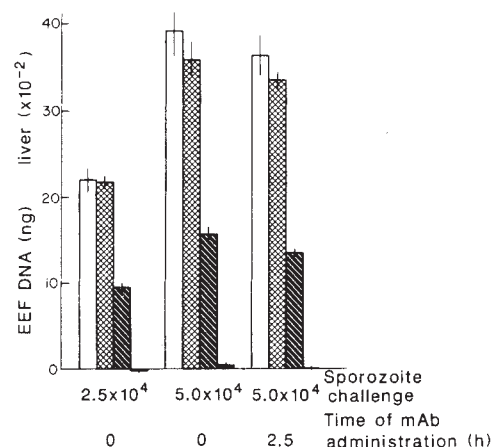


Fig. 1 Endogenous γ IFN inhibits development of EEF in immune hosts. Rats immunized with four doses of 10^5 X-irradiated sporozoites of *P. berghei*, at biweekly intervals, were challenged after three weeks with 2.5×10^4 or 5×10^4 sporozoites. Immediately or 2½ h later, animals received an intravenous inoculation of 1 mg of mAb DB-1, or a control mAb. EEF-DNA levels were determined after 44 h, as described¹⁰. Data are mean values from five animals $\pm$ s.d. □, Naive plus DB-1; ▨, naive plus control mAb; □, immunized plus DB-1; ▨, immunized plus control mAb.

tible to challenge, the MHC Class II restricted T-helper epitopes required for antibody production to the CS protein¹⁷⁻²⁰ are not sufficient to prime T cells for γ IFN mediated protection. CD8⁺ cells recognize MHC Class I restricted epitopes, which may be structurally distinct. Our results suggest that effective human malaria sporozoite vaccines will require the incorporation of synthetic peptide or recombinant protein analogues of parasite B, T-helper, and T-cytotoxic epitopes, capable of inducing high titres of neutralizing antibodies to the CS protein, and of sensitizing T-cell subsets that respond with the production of γ IFN to parasite antigens presented during natural sporozoite challenge.

We are grateful to Drs S. Wasserman, University of Maryland, and Sam Wertheimer, New York University, for help with the statistical analyses. This research was supported by the Agency for International Development, the NIH and the MacArthur Foundation.

Received 17 September; accepted 4 November 1987.

- Nussenzweig, R. & Nussenzweig, V. *Phil. Trans. R. Soc.* **B307**, 117-128 (1984).
- Zavala, F., Cochrane, A., Nardin, E., Nussenzweig, R. & Nussenzweig, V. *J. exp. Med.* **157**, 1947-1957 (1983).
- Potocnjak, P., Yoshida, N., Nussenzweig, R. & Nussenzweig, V. *J. exp. Med.* **151**, 1504-1513 (1980).
- Hollingdale, M., Zavala, F., Nussenzweig, R. & Nussenzweig, V. *J. Immun.* **128**, 1929-1930 (1982).
- Chen, D., Tigelaar, R. & Weinbaum, F. *J. Immun.* **118**, 1322-1327 (1977).
- Ferreira, A. *et al. Science* **232**, 881-884 (1986).
- Maheshwari, R. *et al. Infect. Immun.* **53**, 628-630 (1986).
- Schofield, L., Ferreira, A., Altszuler, R., Nussenzweig, V. & Nussenzweig, R. S. *J. Immun.* **139**, 2020-2025 (1987).
- Van Der Meide, P. *et al. J. gen. Virol.* **67**, 1059 (1986).
- Ferreira, A., Enea, V., Morimoto, T. & Nussenzweig, V. *Molec. Biochem. Parasitol.* **19**, 103-109 (1986).
- Shin, S., Vanderberg, J. & Terzakis, J. *J. Protozool.* **29**(3), 448-454 (1982).
- Hamilton Fairley, N. *Trans. R. Soc. trop. Med. Hyg.* **40**, 621-676 (1947).
- Cobbold, S., Jayasuriya, A., Nash, A., Prospero, T. & Waldmann, H. *Nature* **312**, 548-551 (1984).
- Verhave, J., Strickland, G., Jaffe, H. & Ahmad, A. *J. Immun.* **121**, 1031-1033 (1978).
- Egan, J. *et al. Science* **236**, 453-456 (1987).
- Zavala, F. *et al. J. exp. Med.* **166**, 1591-1596 (1987).
- Good, M. *et al. Science* **235**, 1059-1062 (1976).
- Good, M. *et al. J. exp. Med.* **164**, 655-660 (1986).
- Del Giudice, G. *et al. J. Immun.* **137**, 2952-2955 (1986).
- Togna, A. *et al. J. Immun.* **137**, 2956-2960 (1986).

A novel steroid thyroid hormone receptor-related gene inappropriately expressed in human hepatocellular carcinoma

Hugues de Thé, Agnès Marchio, Pierre Tiollais & Anne Dejean

Unité de Recombinaison et Expression Génétique (INSERM U.163, CNRS UA 271), Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cédex 15, France

We have previously isolated from a human hepatocellular carcinoma a hepatitis B virus integration in a 147-base-pair cellular DNA fragment, similar to steroid- and *c-erb-A*/thyroid-hormone receptor genes¹. We have now cloned the corresponding complementary DNA from a human-liver cDNA library. Nucleotide sequence analysis revealed that the overall structure of the cellular gene, which we have named *hap*, is similar to that of the DNA-binding hormone receptors. That is, it displays two highly conserved regions identified as the putative DNA-binding and hormone-binding domains of the *c-erbA*/steroid receptors²⁻⁹. Six out of seven hepatoma and hepatoma-derived cell-lines express a 2.5-kilobase (kb) *hap* messenger RNA species which is undetectable in normal adult and fetal livers but present in all non-hepatic tissues analysed. The data suggest that the *hap* gene product may be a novel ligand-responsive regulatory protein whose inappropriate expression in liver may relate to the hepatocellular carcinogenesis.

We screened a human-liver cDNA library using as a probe the putative 147-bp cellular exon in which hepatitis B (HBV) integration took place¹. The longest cDNA clone (Fig. 1a), 3 kb in length, was subjected to nucleotide-sequence analysis. The deduced amino-acid sequence encoded by this gene, which we proposed to name *hap* for hepatoma, revealed a long open reading frame of 448 amino acids corresponding to a predicted polypeptide of relative molecular mass M_r 51,000 (51K) (Fig. 1b).

The similarity previously reported between the putative 147-bp cellular exon (bracketed in Fig. 1b) and the *c-erbA*/steroid receptor genes led us to compare the entire predicted amino-acid sequence of the *hap* gene product with the glucocorticoid receptor (GR)^{10,11}, the progesterone receptor (PR)^{6,12}, the oestrogen receptor (ER)^{4,5} and the *hc-erbA*/thyroid hormone receptor^{2,3} (Fig. 2a). The sequence-comparison analysis revealed that the overall organization of the predicted *hap* gene product is very similar to that of the four receptors in that it can be roughly divided into four regions (arbitrarily referred to as A/B, C, D and E¹³). Comparison between the cysteine-rich C region of the predicted *hap* gene product (residues 81–146) with the corresponding region of the four receptors reveals 64% amino-acid identity with *hc-erbA*, 59% with hER, 42% with rPR and 44% with hGR (schematically represented in Fig. 2b). Apart from *c-erbA*, which contains two additional residues, the 66 amino-acid long C region shows a constant length in hER, hGR, rPR and predicted *hap* sequences. Region E (residue 195–448), which is well-conserved but to a lesser extent, is slightly more similar to *hc-erbA* (38%) (Fig. 2b). Different functions have been assigned to the four regions defined in the glucocorticoid and oestrogen receptors^{7,9,14}. By analogy, we speculate that the two highly conserved regions C and E may represent respectively the putative DNA-binding and hormone-binding domains of the predicted *hap* protein.

The ability to express functional receptors *in vitro* from cloned *c-erbA*/steroid receptor genes led us to use *in vitro* translation assay to identify a putative *hap* ligand. Figure 3 shows that the *hap* RNA directed the efficient synthesis of a major protein of 51K, consistent with the size predicted by the amino-acid sequence (lanes a and b). Because *c-erbA* and *hap* are more closely related according to their amino-acid sequence, we first

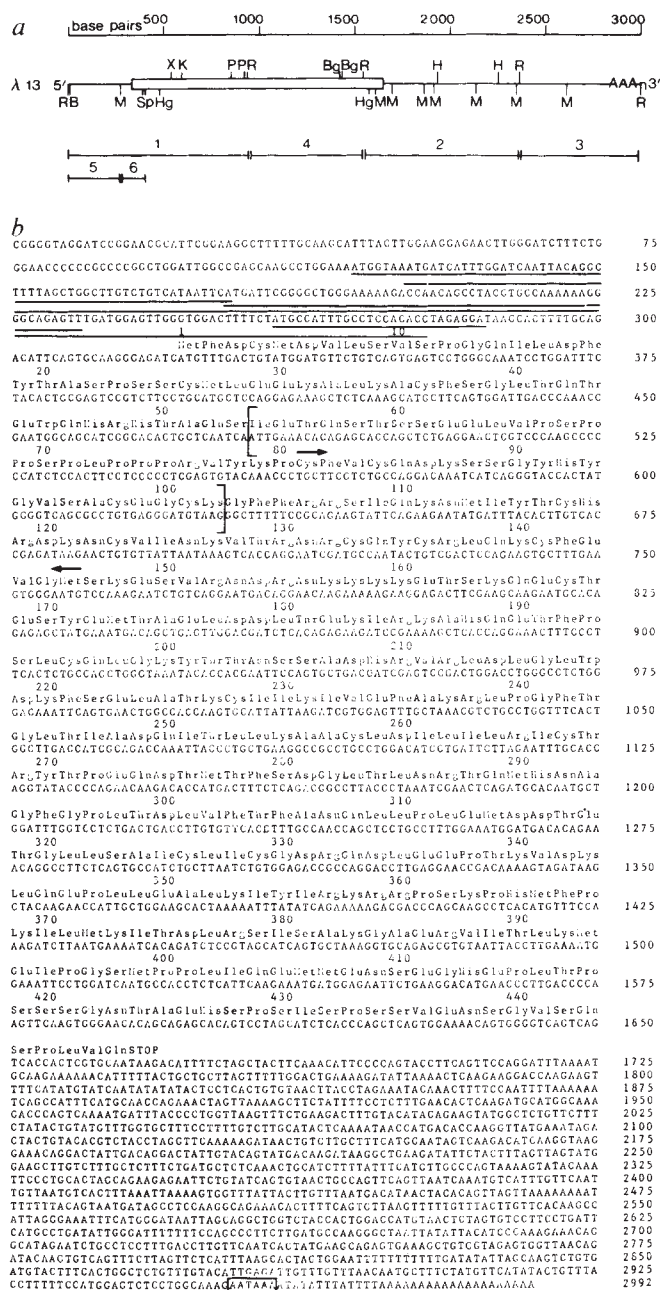


Fig. 1 Restriction map and nucleotide and predicted amino-acid sequence of human liver *hap* cDNA. **a**, The insert of clone λ13, representing a nearly full-length cDNA for the *hap* gene, is represented and noncoding (lines) and coding (boxed portion) sequences are indicated. Restriction sites are: R, *EcoRI*; Bg, *BglII*; M, *MaeI*; X, *XhoI*; K, *KpnI*; P, *PvuII*; B, *BamHI*; H, *HindIII*; Sp, *SphI*; Hg, *Hgi* A1. Probes 1–6 are specific for defined regions in the *hap* cDNA clone. They correspond to *EcoRI* fragments for probes 1–4 and to *EcoRI*, *MaeI* and *MaeI*–*SphI* fragments for probes 5 and 6, respectively. **b**, The nucleotide sequence of the *hap* cDNA is presented in the 5' to 3' orientation. The numbers on the right refer to the position of the nucleotides. Numbers above the deduced translated sequence indicate amino-acid residues. The four short open reading frames in the 5' untranslated region are underlined. Adenosine residues are found at the 3' end of λ13 and the putative polyadenylation-signal site (AATAAA) is boxed. The region similar to the DNA-binding domain of the thyroid/steroid hormone receptors is indicated by horizontal arrows. The exon, previously cloned from a HCC sample genomic DNA library and in which HBV integration took place, is bracketed. Clone λ13 DNA was sonicated and sequenced by the dideoxy-chain termination procedure²². **Methods.** **a**, The cDNA was synthesized using oligo(dT)-primed poly(A)⁺ human liver mRNA, using the method of Gubler and Hoffman²¹ (C. de Taisne, unpublished data). cDNAs of a mean size of 3 kb were selected and cloned after the addition of *EcoRI* linkers into λNM-1149. Four positive 3' co-terminal clones were isolated from the 2×10^6 plaques screened. We used the 350-bp *EcoRI*–*EcoRI* genomic fragment previously referred to as MNT¹ as a probe.

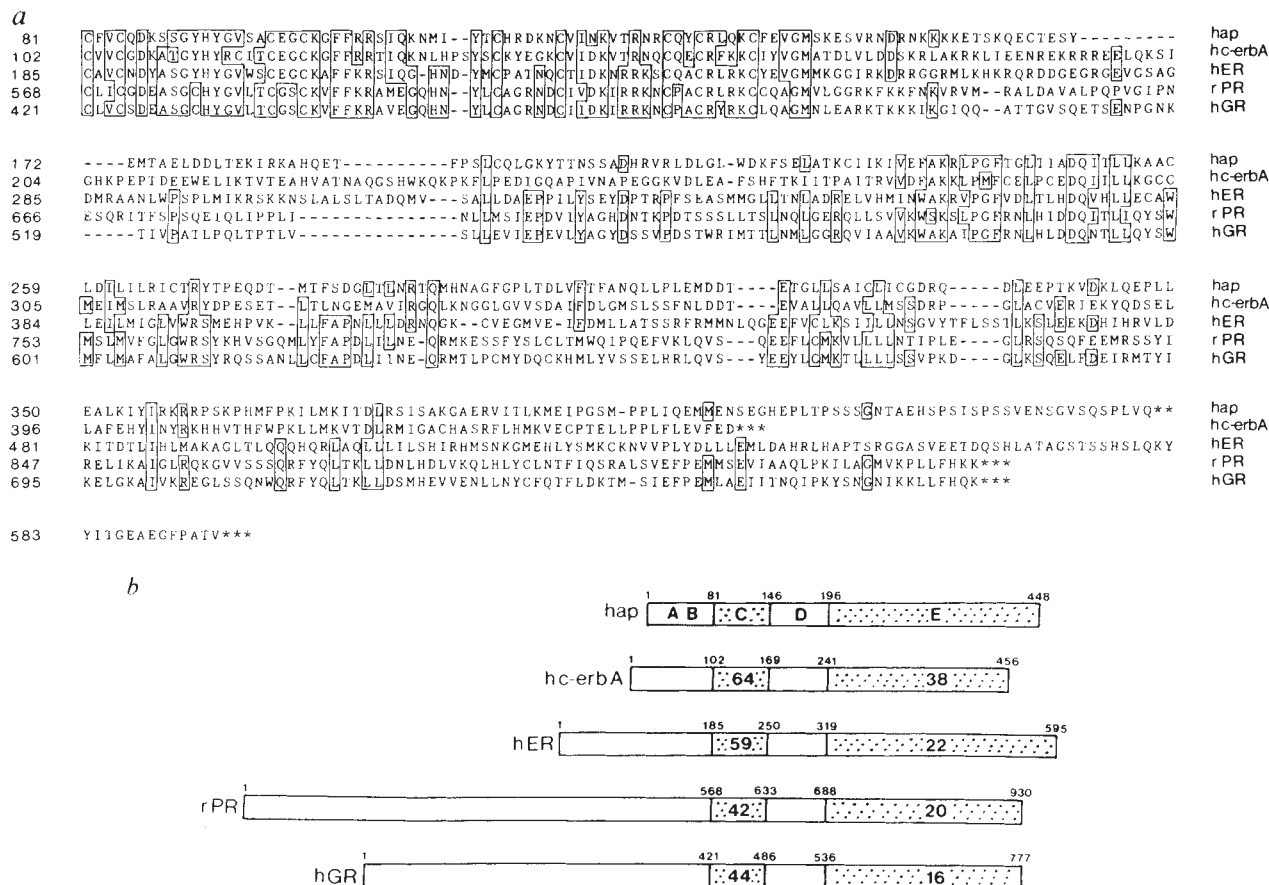


Fig. 2 Alignment of the *hap* translated amino-acid sequence and the thyroid/steroid-hormone receptors. **a**, The five sequences: *hap* product, the human placenta c-erbA protein (hc-erbA)³, the human oestrogen receptor (hER)⁴, the rabbit progesterone receptor (rPR)¹² and the human glucocorticoid receptor (hGR)¹⁰ were aligned after a computer alignment by pairs²³. The amino-terminal ends of the proteins are not represented as no significant similarity was observed in this region. A minimal number of gaps (—) were introduced in the alignment. Amino-acid residues which matched in at least three of the polypeptides are boxed. The one letter code for amino acids is used. **b**, Schematic alignment of the five proteins. The divisions of the thyroid/steroid-hormone receptors (arbitrarily named A/B, C, D, E¹⁶) are schematically represented in the *hap* protein. The two highly conserved regions, identified as the putative DNA-binding (region C) and hormone-binding (region E) domains of the receptors, are shown as stippled blocks. The numbers refer to the position of amino-acid residues. The sequences of each of the hc-erbB product, hER, rPR and hGR receptors are compared with the *hap* protein. The numbers present in the stippled blocks correspond to the percentage of similarity between *hap* protein and each of the receptors in the two highly conserved regions C and E. No significant similarity was observed when comparing the A/B and D regions. The empty blocks correspond to the non-conserved A/B and D regions.

tested [¹²⁵I]-T3 (triiodothyronine), reverse T3 (3,3',5'-triiodo-L-thyronine) and T4 (thyroxine) for their binding with the polypeptide translated from the *hap* gene. No specific fixation with any of those three thyroid hormones could be detected. The results were also negative when we repeated the experiment with tritiated retinol, retinoic acid and testosterone, which represent three putative ligands for *hap* whose receptors have not yet been cloned.

To study the expression of the *hap* gene, we performed Northern-blot analysis using the cDNA insert as a probe and poly(A)⁺ RNA extracted from various human tissues and cell lines (Fig. 4A). Two RNA species of 3 kb and 2.5 kb (the size of this smaller mRNA is slightly variable from one organ to another) were expressed at low abundance in ovary (lane a), uterus (lane b), HBL100 mammary cells (lane c), adult and fetal spleen (lane d and e respectively), K562 and HL60 haematopoietic cell lines (lanes f and g). Surprisingly, expression was increased about tenfold in prostatic adenoma (lane h) and kidney (lane i). By contrast, a single mRNA of 3,000 nucleotides, expressed at low levels, was present in poly(A)⁺ RNA from adult and fetal liver tissues (lanes j and k). The smaller 2.5 kb mRNA was undetectable, even after long exposure, in three adult and two fetal human livers analysed (Fig. 4B, lane a). The over-expression of two mRNA species in prostate and kidney, as well as the

presence of a single mRNA expressed at low level in adult and fetal livers show that *hap* expression is differentially regulated in those organs. This tissue-specific expression provides some indication that prostate and kidney as well as liver could be key tissues and that *hap* functions in those cell types may differ.

Because we first identified *hap* in a human primary liver cancer, poly(A)⁺ RNA from seven hepatoma and hepatoma-derived cell lines were analysed by Northern blotting. In addition to the 3 kb-long mRNA found in normal adult and fetal liver, an additional 2.5 kb RNA species was observed, in equal or even greater amount, in three out of four human hepatocellular carcinomas (HCC) (Fig. 4B, lanes b, c and d) and in the PLC/PRF/5, HEPG2 and HEP3B hepatoma cell lines (lanes e, f and g). Also the two transcripts were strikingly overexpressed, at least tenfold, in the PLC/PRF/5 cells. After Southern-blotting analysis, no rearrangement of the *hap* gene was visible at the genomic level (data not shown).

To determine more precisely the structure of the smaller 2.5-kb mRNA present in hepatocarcinoma, we performed Northern blots using poly(A)⁺ RNA extracted from the PLC/PRF/5 hepatoma cell line and the probes 1–6 specific for defined regions in the *hap* cDNA insert (described in Fig. 1a). Although the 3 kb transcript hybridized to all six probes, the smaller mRNA species was detected by probes 1, 2, 3, 4 and 6 but not by probe

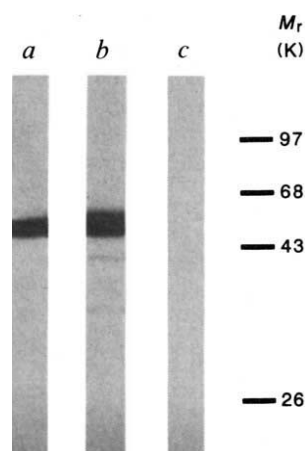


Fig. 3 Synthesis of *hap* polypeptide *in vitro*. Lane *a*, pCOD 20 (sense RNA, 70 ng). Lane *b*, pCOD 20 (140 ng). Lane *c*, pCOD 14 (antisense RNA, 140 ng).

Methods. The [35 S]methionine-labelled products were separated on a 12% SDS-polyacrylamide gel which was fluorographed. The *MaeI*-*MaeI* fragment, extending from the first to the third *MaeI* site in the cDNA insert sequence and containing the complete coding region (Fig. 1*a*) was subcloned into pTZ18 (Pharmacia). Capped mRNA was generated using 5 μ g of *HindIII* linearized DNA, 0.5 mM rNTP, 25 mM dithiothreitol (DTT), 100 U RNasin (Promega), 50 U T7 Pol (Genofit) in 40 mM Tris pH8, 8 mM $MgCl_2$, 2 mM spermidine, 50 mM NaCl, in 100 μ l at 37 $^{\circ}$ C. Capping was performed by omitting GTP and adding CAP (m⁷G(5')ppp(5')G) (Pharmacia) for the first 15 min of the reaction. Translation was performed using rabbit reticulocyte lysate (Amersham) using 40 μ l of lysate for 2.5 μ g of capped RNA. The thyroid hormones binding assays included 5 μ l lysate in (0.25 M sucrose, 0.25 KCl, 20 mM Tris (pH 7.5), 1 mM $MgCl_2$, 2 mM EDTA, 5 mM DTT) with 1 nM [125 I]T₄, [125 I]T₃ or [125 I]rT₃ (specific activity: T₄, rT₃, 1,400 mCi mg⁻¹ (Amersham); T₃, 3,000 mCi mg (NEN)). After incubation at 0 $^{\circ}$ C, free ligand was separated from bound ligand by filtration (Millipore HAWP 02500). For testosterone, retinol and retinoic acid 10 μ l of lysate were added to 45 μ l of 20 mM Tris, pH 7.3, 1 mM EDTA, 50 mM NaCl, 2 mM β mercaptoethanol and 5 nM testosterone, 400 nM retinol or 15 nM retinoic acid (81 Ci mmol⁻¹; 60 Ci mmol⁻¹; 46 Ci mmol⁻¹, Amersham). After an overnight incubation at 0 $^{\circ}$ C free ligand was separated from bound ligand by Dextran-coated charcoal extraction. All experiments were performed in duplicates and parallel experiments were performed with 100-fold excess corresponding cold hormone. As a positive control we detected binding of T₃ with nuclear extracts from HeLa cells.

5 which anneals to the complete 5' untranslated region (data not shown). This result shows that the two *hap* transcripts differ, at least, in their 5' untranslated leader sequence that has been shown, in some cases, to play a role in determining mRNA translational efficiency¹⁵⁻¹⁷.

The *hap* gene product displays two regions similar to the DNA- and hormone-binding domains of previously cloned receptors, and is probably a novel ligand-dependent transcriptional activator. It seems to differ too much from previously cloned hormone receptors to be a variant of one of them, furthermore the *in vitro* translated 51K *hap* polypeptide failed to bind all ligands tested. Although one cannot exclude the possibility that that *hap* product could be hormone independent, we believe that *hap* encodes a receptor for a presently unknown circulating or intracellular ligand. Because *hap* is both linked to the steroid-receptor gene by its shorter C domain and to the thyroid-hormone-receptor genes by its greater similarity with *c-erb-A* in the E region, we suggest that the *hap* ligand may belong to a different hormone family.

The known steroid/thyroid-hormone receptors control the expression of target genes that are crucial for cell growth and

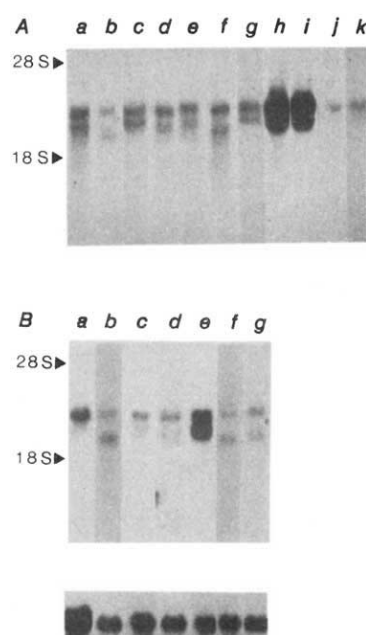


Fig. 4 Distribution of *hap* gene transcripts. Northern blot analysis were performed with poly(A)⁺ RNAs (15 μ g per lane) extracted from different human organs and cell lines. A control hybridization with a mouse β actin cDNA probe is shown below the hybridizations. *A*, *Hap* mRNA in different tissues. Lane *a*, ovary; lane *b*, uterus; lane *c*, HBL100 mammary cells; lane *d*, adult spleen; lane *e*, 18 weeks fetal spleen; lane *f*, K562; lane *g*, HL60 haematopoietic cell lines; lane *h*, prostatic adenoma; lane *i*, kidney; lane *j*, adult liver; and lane *k*, 18 weeks fetal liver. Lanes *a*-*k* correspond to a one day exposure. *B*, *Hap* mRNA in HCC and HCC-derived cell lines. Lane *a*, normal liver (four days autoradiography); lanes *b*, *c* and *d*: three HCC samples (lane *b*, patient Ca; lane *c*, patient Mo; lane *d*, patient TC1); lanes *e*, *f* and *g*: three HCC-derived cell lines (lane *e*, PLC/PRF/5; lane *f*, HEPG2; lane *g*, HEP 3B).

The lanes *b*-*g* correspond to a one day exposure.

differentiation¹⁸. An altered and/or deregulated receptor could participate in the cell transformation¹⁹. Indeed the avian *v-erb-A* oncogene, whose product is defective in binding the thyroid hormone², potentiates the erythroblast transformant effects of *v-erb-B* and other oncogenes of the *src* family²⁰. Northern-blot analysis of human HCCs and hepatoma cell-lines showed almost constant alterations in *hap* transcription. There are two possible explanations for this result. The smaller mRNA species could simply be overexpressed as a consequence of the hepatocellular dedifferentiation, but the inability to detect a 2.5-kb transcript in fetal livers does not seem to favour this hypothesis. On the contrary, the presence of the smaller transcript may have preceded the tumorigenic events and therefore reflect a preneoplastic state. The presence of an inappropriately expressed *hap* protein, normally absent from normal hepatocytes, may have directly participated to the hepatocellular transformation.

We thank Drs Marc Alizon, Pierre Sonigo and Jean Weissenbach for advice and discussion, Dr Charles de Taisne (Institut Mérieux) for providing human liver cDNA library, Drs Christian Br  chet, Michelle Hadchouel, Fran  oise M  ller and Bernard Hamelin for providing tissue samples, Dr G  rard Goubin for the HBL100 mammary cell line and Dr Pierre Tambourin for providing poly(A)⁺ from haematopoietic cell lines, Drs Martin Petkovich for help in expression and in hormone-binding assays, Dr Michael Emerman for revising the manuscript and Louise-Marie Da for typing this manuscript. This work was supported by grants from the National Institute of Health and Ligue Nationale Fran  aise contre le Cancer. The

sequence data in this publication have been submitted to the EMBL/GenBank Libraries under the accession number Y00291. *Note added in proof:* A retinoic acid receptor has been recently identified (Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. (*Nature* **330**, 444–450; 1987) and Giguere, V., Ong, E., Segui, P. & Evans, R. M., see pages 624–629, this issue). The *hap* gene is different from the *hRAR* cloned by these two groups as shown by its different chromosomal localization (the *hap* gene maps to human chromosome 3(1)) and by their divergent nucleotide sequence. However, the very high amino-acid sequence homology between *hap* and the *hRAR* cloned by Petkovich *et al.* (90% identity in the E region) strongly suggest that *hap* product may be a retinoid responsive transcriptional activator.

Received 30 September; accepted 16 November 1987.

- Dejean, A., Bougueret, L., Grzeschik, K. H. & Tiollais, P. *Nature* **322**, 70–72 (1986).
- Sap, J. *et al.* *Nature* **324**, 635–640 (1986).
- Weinberger, C. *et al.* *Nature* **324**, 641–646 (1986).
- Green, S. *et al.* *Nature* **320**, 134–139 (1986).
- Greene, G. L. *et al.* *Science* **231**, 1150–1154 (1986).
- Conneely, O. M. *et al.* *Science* **233**, 767–770 (1986).
- Kumar, V., Green, S., Staub, A. & Chambon, P. *EMBO J.* **5**, 2231–2236 (1986).
- Hollenberg, S. M., Giguere, V., Segui, P. & Evans, R. M. *Cell* **49**, 39–46 (1987).
- Miesfeld, R., Godowski, P. J., Maler, B. A. & Yamamoto, K. R. *Science* **236**, 423–427 (1987).
- Hollenberg, S. M. *et al.* *Nature* **318**, 635–641 (1985).
- Miesfeld, R. *et al.* *Cell* **46**, 389–399 (1986).
- Loosfelt, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 9045–9049 (1986).
- Krust, A. *et al.* *EMBO J.* **5**, 891–897 (1986).
- Giguere, V., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. *Cell* **46**, 645–652 (1986).
- Sonenberg, N., Guertin, D. & Cleveland, D. *Cell* **27**, 563–572 (1981).
- McGarry, T. J. & Lindquist, S. *Cell* **42**, 903–911 (1985).
- Jobling, S. A. & Gehrke, L. *Nature* **325**, 622–625 (1987).
- Yamamoto, K. R. *A Rev. Genet.* **19**, 209–252 (1985).
- Green, S. & Chambon, P. *Nature* **324**, 615–617 (1986).
- Kahn, P. *et al.* *Cell* **45**, 349–356 (1986).
- Gubler, U. & Hoffman, B. J. *Gene* **25**, 263–269 (1983).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726–730 (1983).

Transcription in yeast activated by a putative amphipathic α helix linked to a DNA binding unit

Edward Giniger* & Mark Ptashne

Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

Gene activation by a DNA-binding regulatory protein in yeast requires the protein to have two components: one to recognize a specific DNA sequence and a second, the 'activating region', to interact with a general transcription factor or perhaps with RNA polymerase^{1,2}. The activating regions that have been characterized are acidic^{3,4}, and mutational analysis of one indicates that this acidity is important for activity⁶. Here we report the design of an artificial protein bearing a novel 15-amino acid peptide linked to a DNA binding fragment of the yeast regulatory protein GAL4 (refs 7–10). The synthetic peptide is acidic and should it form an α -helix, that helix would be amphipathic, having one hydrophilic face bearing the acidic residues, and one hydrophobic face¹¹. When expressed in yeast, the artificial protein bearing this peptide efficiently activates the *GAL1* gene which is ordinarily activated by GAL4 (refs 12, 13). An otherwise identical protein with the novel 15 amino acids in a scrambled order, and which is thus unable to form an amphipathic structure, does not activate *GAL1* transcription.

GAL4 binds four related sites in the *GAL* upstream activating sequence (UAS_G) and stimulates transcription of the adjacent *GAL1* and *GAL10* genes^{8–10,14}. A protein fragment consisting of the amino-terminal 147 amino acids of GAL4 binds

specifically to DNA *in vivo* but does not activate transcription^{4,10}. The carboxy terminus of GAL4 contains two activating regions, regions I and II, either of which stimulates transcription when linked to an appropriate DNA-binding unit⁴. The yeast regulatory protein GCN4 also bears an activating region which is separable from its DNA-binding region³. An array of peptides that function as activating regions have been identified by fusing random *Escherichia coli* DNA fragments to sequences encoding the DNA-binding unit of GAL4, and one of these peptides also activates transcription when fused to the DNA-binding domain of the bacterial repressor *lexA* (ref. 5). All of the known activating regions are acidic, but they have no obvious sequence homology. Many of these sequences, however, and particularly the *E. coli*-encoded activating peptides, contain amino-acid sequences which, should they fold into α helices, would form short, negatively-charged, amphipathic α helices with an acidic face and a hydrophobic face. In most cases, the third face of the putative helix is uncharged but polar, though in some cases this face also carries one or more acidic residues.

To test whether an acidic, amphipathic α helix could stimulate transcription, we constructed genes for two artificial proteins (see Fig. 1 and legend). Both genes code for the amino-terminal 147-amino acid, DNA-binding fragment of GAL4, followed by a short polylinker. One gene then ends with a sequence coding for the peptide NH₂-(Glu-Leu-Gln)₃-Ala-(Leu)₂-(Gln)₃-COOH which could form a four-turn, negatively-charged, amphipathic α helix. The other gene terminates with a sequence coding for the peptide NH₂-(Leu)₃-(Glu-Gln)₃-Ala-(Leu)₂-(Gln)₃-COOH, which would not form an amphipathic structure if it folded into an α helix. Both artificial proteins were expressed in yeast, and assayed for activation of a *GAL1-lacZ* fusion gene^{14,15}.

In yeast, the protein bearing a putative amphipathic α helix stimulates transcription to ~20% of the level directed by wild-type GAL4, whereas the protein bearing a scrambled version of the helix does not detectably stimulate transcription (Fig. 1). Moreover, a protein bearing a peptide with the same amino acids as those described above, but in a different scrambled order, was also essentially inactive, while a protein bearing a putative amphipathic helix with the polar face of glutamine residues replaced by a second hydrophobic face of valine residues also activates transcription, albeit tenfold less efficiently than the most active synthetic protein (see Fig. 1 legend). Expression of a protein fragment which comprises the first 147 amino acids of GAL4 followed by polylinker-encoded amino acids, but has no artificial helix, does not activate transcription. It should be noted that we have no independent evidence that our artificial peptides fold into α helices.

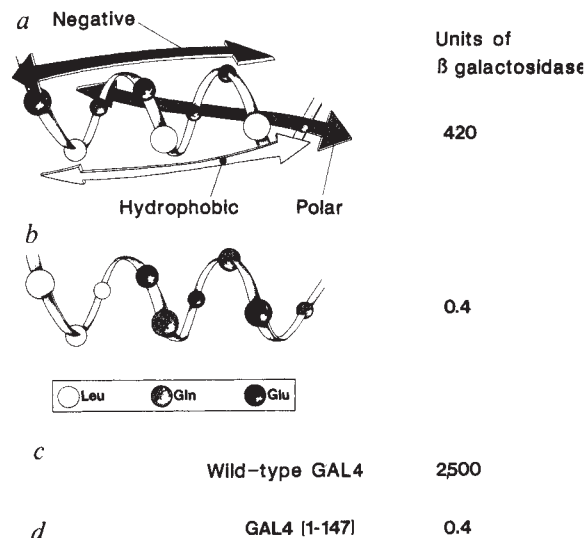
The interpretation of our results turns on the observation that the scrambled putative helix fails to activate transcription, so it is crucial to demonstrate that the inactive artificial protein binds UAS_G *in vivo*. The transcriptionally active and inactive artificial proteins have indistinguishable relative molecular masses, as assayed by immunoblotting of polyacrylamide gels of cell extracts containing these proteins (Fig. 2); moreover both proteins are larger than the purified 147-amino acid DNA-binding fragment of GAL4. Footprinting UAS_G *in vivo* (Fig. 3) shows that the methylation protection pattern in cells expressing the inactive artificial protein is the same as that seen in cells containing wild-type GAL4⁹. Therefore the artificial protein bearing a scrambled helix binds UAS_G in cells in which it does not activate transcription. We conclude that the artificial protein which fails to activate transcription is produced and is stable, and binds UAS_G efficiently *in vivo*. These data support the idea that the amphipathic nature of our putative helices is necessary for activity.

We do not know how commonly amphipathic α helices function as transcriptional activating regions. The regions of bacteriophage λ and 434 repressors that stimulate transcription are known from X-ray crystallography to be negatively-charged, amphipathic α helices, and it has been argued that these function

* Present address: Howard Hughes Medical Institute, Department of Physiology, University of California, San Francisco, California 94143, USA.

Fig. 1 Activation of transcription by artificial proteins. Proteins to be assayed for transcriptional activation were expressed in yeast which contained a *GAL1-lacZ* fusion gene but lacked GAL4, and β -galactosidase activity was measured. *a*, Activation of transcription by a protein bearing a putative amphipathic α helix (encoded by pEG50); *b*, activation by a protein bearing a scrambled peptide which could not form an amphipathic structure (encoded by pEG52); *c*, activation by wild-type GAL4; *d*, activation by the DNA-binding unit of GAL4 attached to polylinker-encoded amino acids but no novel peptide. Transcription induced by the protein bearing a putative amphipathic α helix initiated at the correct nucleotide, as assayed by primer extension, and RNA levels were consistent with those expected based on the β -galactosidase measurements (data not shown). Moreover, activation of transcription by this protein required the presence of a GAL4 binding site upstream of the *GAL1-lacZ* fusion gene (data not shown). Note that all our synthetic peptides end with the sequence -Ala-(Leu)₂-(Gln)₃-COOH. In most of the *E. coli*-encoded activating peptides⁵ the hydrophobic face of the putative amphipathic helix extended one helical turn beyond the final acidic residue of the negatively-charged face, and bore two hydrophobic residues in the final amphipathic helical turn. Hence we appended -Ala-(Leu)₂ to the peptides depicted in the figure. The three final glutamine residues were included to stabilize the carboxy-terminal end of the putative helix. Another artificial protein bearing a peptide that contained the same amino acids as the peptides in *a* and *b*, but in a different scrambled order (NH₂-(Glu-Leu)₃-(Gln)₃-Ala-(Leu)₂-(Gln)₃-COOH), was essentially inactive in directing transcription (7 units of β -galactosidase). In contrast, a protein bearing a putative amphipathic α helix with a slightly different amino-acid composition did stimulate correctly initiated transcription, inducing the synthesis of 35 units of β galactosidase. This helix, NH₂-(Glu-Leu-Val)₃-Ala-(Leu)₂-(Gln)₃-COOH, was more distantly related to the proposed activator motif than was the amphipathic helix in *a*, as it had a second hydrophobic face containing valine residues in place of the polar face of glutamine residues. The mean helical hydrophobic moments²⁶ of the acidic portions of the peptides pictured in *a* and *b* are, respectively, 0.537 and 0.324, and that of the second scrambled peptide is 0.227. The first of these values is typical for a "surface-seeking" α -helix such as a calmodulin-binding peptide, whereas the latter two values are in the range typical for α -helices buried in globular proteins²⁷. Note that calculation of the mean helical hydrophobic moment of the transcriptionally active peptide as a single unit obscures its local amphipathicity; the hydrophobic and hydrophilic surfaces wrap partway around the putative helix and hence tend to cancel in net amphipathicity over the length of the entire helix. Such wrapping is commonly found for contacted surfaces of α helices from natural proteins and is thought to arise from the way in which α helices are packed against one another in globular proteins²⁸.

Methods. Plasmids encoding artificial proteins were derived from pMA424 (ref. 5), which bears the first 147 codons of GAL4, followed by a polylinker, transcribed from the yeast *ADH1* promoter. Synthetic oligonucleotides encoding the novel peptides (Harvard Biolabs Microchemistry Facility) were cloned by standard methods¹⁶ into the *Bam*HI site in this polylinker. Thus, the artificial proteins had the same sequence as GAL4 up through amino acid 147, then the sequence PEFPGI, followed by the peptides described. The plasmid bears a *HIS3* selectable marker and the 2 μ plasmid ARS, as well as *amp*^R and the origin of replication from pBR322. The plasmid expressing wild-type GAL4, pMA210 (ref. 5), is identical to pMA424 outside the GAL4 coding sequence. The host strain for all experiments was GGY1 (*ura3-52; leu2-2,112; his3- Δ 200 Δ gal14*) (ref. 6), into which a single copy of a *GAL1-lacZ* fusion was integrated at the *URA3* locus. Plasmids were transformed into yeast cells treated with lithium acetate¹⁷ and grown in minimal media¹⁸ which lacked histidine and contained 2% galactose, 5% glycerol and 2% potassium lactate at pH5.5 as carbon source. Alternatively, were pre-grown in his⁻ media containing glucose and then back-diluted into his⁻ media containing galactose glycerol and lactate. Exponentially growing cells were assayed for β -galactosidase activity by standard procedures¹⁴; standard error was less than 20% when cultures were assayed in triplicate. Primer-extension analysis of RNA was as described⁴.



by contact with RNA polymerase²²⁻²⁵. Activating region I of GAL4 and the activating region of GCN4 include sequences consistent with formation of an amphipathic helix. In contrast, activating region II of GAL4, and one of the *E. coli*-encoded activating peptides, B17 (ref. 5), apparently cannot form amphipathic α helices. We note, however, that the spacing between

acidic residues within these two sequences is the same, and perhaps they are examples of some specific alternative activation structure.

We thank members of our laboratory, particularly A. Brand, M. Carey, G. Gill, H. Himmelfarb, A. Hochschild, J. Koudelka, and J. Ma, as well as E. Ostrander, for discussion and critical

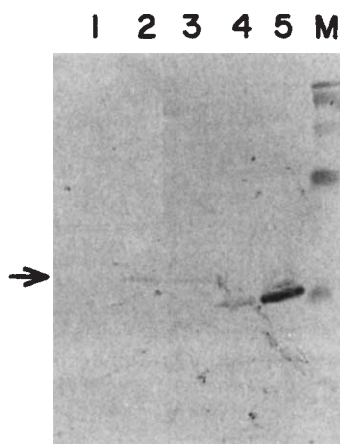


Fig. 2 Immunoblot of cell extracts containing artificial proteins. Extracts of yeast cells expressing artificial proteins were analysed by polyacrylamide-gel electrophoresis and western blotting, using antiserum raised against the purified DNA-binding unit of GAL4 as probe. Arrow indicates the position of the two artificial proteins. Lane 1, no GAL4; lane 2, protein bearing a scrambled helix (encoded by pEG52); lane 3, protein bearing a transcriptionally active putative amphipathic helix (encoded by pEG50); lane 4, 147-amino acid DNA-binding unit of GAL4, purified from *E. coli* (10 ng); lane 5, DNA-binding unit of GAL4 (100 ng); M; pre-stained protein molecular weight markers (BioRad). The marker protein that comigrates with the DNA-binding unit of GAL4 is lysozyme (*M*_r 17,000). Note that the protein bearing a scrambled helix was as abundant as the protein bearing a transcriptionally active helix. We have not sequenced our two artificial proteins, so it is conceivable that there has been a small amount of proteolysis of one or both. The band migrating more quickly than the DNA binding unit of GAL4 is an uncharacterized yeast protein that cross-reacts with our anti-GAL4 antiserum.

Methods. Exponentially growing yeast cultures (5 ml) were centrifuged at 5,000 r.p.m. for 5 min, and the cell pellets resuspended in 200 μ l sample buffer for SDS-PAGE, vortexed with glass beads for 30 s, then boiled for 4 min and fractionated on an SDS/urea polyacrylamide gel¹⁹. Western transfer and probing with antiserum were as described²⁰; primary antibody was visualized with goat anti-rabbit IgG coupled to horse radish peroxidase (BioRad).

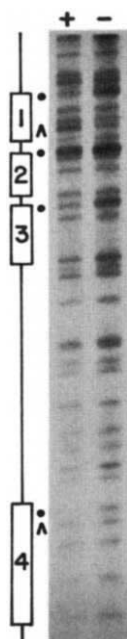


Fig. 3 Methylation footprint *in vivo* of the artificial protein bearing a scrambled helix (encoded by pEG52). Cells that expressed (+) or did not express (−) the transcriptionally inactive artificial protein were treated with dimethylsulphate, the DNA isolated, and guanines of UAS_G visualized by genomic sequencing^{9,21}. The positions of the GAL4-binding sites are indicated alongside the autoradiograph. Residues that were protected from methylation by GAL4 binding (●); residues whose reactivity was enhanced by GAL4 binding, (△). The top strand of UAS_G (ref. 15) is shown. Note that the incompletely protected band in site 2 arose from compression in the gel of three consecutive guanine residues; only one of these is normally protected by GAL4. The apparent methylation effects in this autoradiograph that are unrelated to GAL4 binding were not reproducible and are not highlighted. Methylation of cells, purification of DNA and genomic sequencing were as described^{9,21}.

reading of the manuscript, and E. Burkhardt for artwork. We thank W. DeGrado for advice about designing amphipathic helices, G. Gill and J. Ma for strains and plasmids, H. Kakidani for purified GAL4 DNA-binding domain, and S. Chapman for anti-GAL4 antiserum. We acknowledge the support of the American Cancer Society.

Received 24 September; accepted 2 November 1987.

1. Ptashne, M. *Nature* **322**, 697–701 (1986).
2. Struhl, K. *Cell* **49**, 295–297 (1987).
3. Hope, I. A. & Struhl, K. *Cell* **46**, 885–894 (1986).
4. Ma, J. & Ptashne, M. *Cell* **48**, 847–853 (1987).
5. Ma, J. & Ptashne, M. *Cell* **51**, 113–119 (1987).
6. Gill, G. & Ptashne, M. *Cell* **51**, 121–126 (1987).
7. Klar, A. & Halvorsen, H. *Molec. gen. Genet.* **135**, 203–212 (1974).
8. Bram, R. & Kornberg, R. D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 43–47 (1985).
9. Giniger, E., Varnum, S. M. & Ptashne, M. *Cell* **40**, 767–774 (1985).
10. Keegan, L., Gill, G. & Ptashne, M. *Science* **231**, 699–704 (1986).
11. O'Neil, K. T., Wolfe, H. R. Jr, Erickson-Viitanen, S. & DeGrado, W. F. *Science* **236**, 1454–1456 (1987).
12. Hopper, J., Broach, J. & Rowe, L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2878–2882 (1978).
13. St. John, T. & Davis, R. J. *molec. Biol.* **152**, 285–315 (1981).
14. West, R. Jr, Yocum, R. & Ptashne, M. *Molec. cell. Biol.* **4**, 2467–2478 (1984).
15. Yocum, R., Hanley, S., West, R. Jr & Ptashne, M. *Molec. cell. Biol.* **4**, 1985–1998 (1984).
16. Maniatis, T., Fritsch, E. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor, New York, 1982).
17. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. *J. Bact.* **153**, 163–168 (1983).
18. Sherman, F., Fink, G. & Hicks, J. *Methods in Yeast Genetics* (Cold Spring Harbor, New York, 1983).
19. Wharton, R. P. thesis, Harvard Univ. (1986).
20. Silver, P. A., Keegan, L. P. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5951–5955 (1984).
21. Church, G. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991–1995 (1984).
22. Pabo, C. O. & Lewis, M. *Nature* **298**, 443–447 (1982).
23. Hochschild, A., Irwin, N. & Ptashne, M. *Cell* **32**, 319–325 (1983).
24. Anderson, J. A., Ptashne, M. & Harrison, S. C. *Nature* **316**, 596–601 (1985).
25. Bushman, F. D. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9353–9357 (1986).
26. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. J. *molec. Biol.* **179**, 125–142 (1984).
27. Erickson-Viitanen, S. & DeGrado, W. F. *Meth. Enzym.* **139**, 455–478 (1987).
28. Chothia, C., Levitt, M. & Richardson, D. J. *molec. Biol.* **145**, 215–250 (1981).

Isolated flagellar outer arm dynein translocates brain microtubules *in vitro*

Bryce M. Paschal, Stephen M. King, Anthony G. Moss, Christine A. Collins, Richard B. Vallee & George B. Witman

Cell Biology Group, Worcester Foundation for Experimental Biology, 222 Maple Avenue, Shrewsbury, Massachusetts 01545, USA

The inner and outer arms of the flagellar axoneme generate the forces needed for flagellar movement; these arms contain ATPases called dyneins. To date, there has been no method for studying the mechanochemical transducing activity of isolated dyneins. Recently, it was found that the brain microtubule-associated protein (MAP) 1C is a microtubule-activated ATPase with the structural and force-producing properties of dynein^{1,2}. MAP 1C translocates microtubules in an *in vitro* gliding assay, suggesting that such an assay could also be used with axonemal dyneins. Here, we demonstrate that outer-arm dynein isolated from sea urchin (*Strongylocentrotus purpuratus*) sperm and adsorbed to a glass coverslip can translocate calf-brain microtubules along the surface of the coverslip. Our results conclusively demonstrate that outer-arm dynein by itself is capable of generating shearing forces. The ability to examine the force-generating properties of flagellar dynein *in vitro* should greatly facilitate studies of the mechanism of action of this important mechanochemical transducer.

In the intact flagellum, dynein arms produce shearing forces that result in sliding between adjacent doublet microtubules. This sliding can be demonstrated directly by limited digestion of demembrated axonemes with a protease in the presence of MgATP^{2−}; this causes the doublets to slide actively past one another, resulting in disintegration of the axoneme³. Axonemal disintegration has been used extensively to study the mechanism of flagellar motility^{4–8}; even so, the uncertain effects of protease digestion and the complexity of structural interactions in the axoneme limits the information that can be obtained in this way. Dissection of the mechanism of force generation by dynein would be much easier if the force-producing properties of the arm could be studied *in vitro*, as has been the case with myosin^{9,10}.

Outer-arm dynein was extracted from flagellar axonemes of *S. purpuratus*¹¹ and purified by sucrose density gradient centrifugation (Fig. 1a). Virtually all the ATPase activity and most of the protein sedimented in a single peak. Electrophoretic analysis of the gradient fractions confirmed that this peak contained the α - and β -heavy chains and the three intermediate chains corresponding to sea urchin outer-arm dynein¹¹ (Fig. 1b).

To examine the ability of the purified dynein to translocate microtubules, aliquots of the sucrose gradient fractions were applied to coverslips and the protein allowed to adsorb to the glass surface. Taxol-stabilized microtubules assembled from purified calf-brain tubulin were then added together with MgATP^{2−}, and the preparation examined by video-enhanced differential interference contrast microscopy^{1,15–17}.

In the presence of outer-arm dynein, microtubules attach to the coverslip and actively glide over its surface (Fig. 2a). The distribution of microtubule-translocating activity in the sucrose gradient parallels the distribution of outer-arm dynein ATPase activity (Fig. 1a, b), confirming that the translocating activity is due to dynein. In peak fractions, most of the microtubules in the plane of the coverslip are motile. The microtubules move very rapidly in a unidirectional manner; some stop and then start again in the same direction. Shorter microtubules exhibit considerable side-to-side movement as they progress (Fig. 2b), suggesting that they are bound to the coverslip by relatively few crossbridges. The microtubules appear to be less adherent to

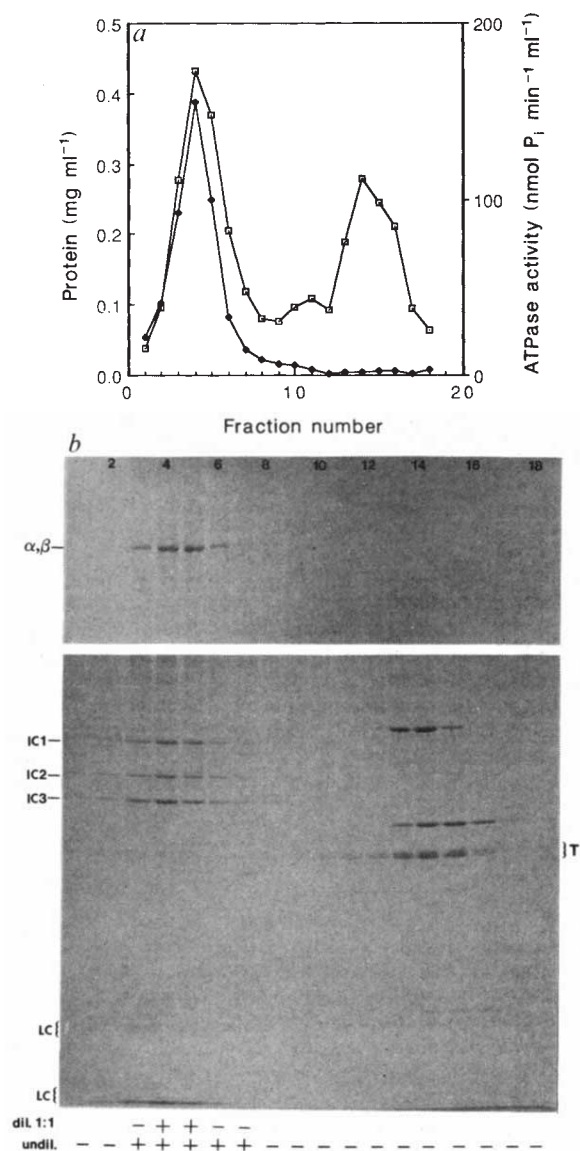


Fig. 1 Density gradient centrifugation of outer-arm dynein. Dynein was extracted from isolated axonemes of *S. purpuratus* sperm using 0.6 M NaCl¹¹ and sedimented through a 5–20% sucrose density gradient in 20 mM Tris-HCl, pH 7.6, 50 mM KCl, 5 mM MgSO₄, 0.5 mM EDTA and 1 mM dithiothreitol (DTT)¹ for 18 h at 107,000g. Fractions of 0.6 ml were collected from the bottom of the gradient. *a*, Protein concentration¹² (□) and Mg²⁺ATPase activity¹³ (●); *b*, electrophoretic analysis¹⁴. In *b*, the upper panel shows the polypeptides present in the high *M_r* region of a 3–5% acrylamide gradient gel; the lower panel shows proteins of intermediate and low *M_r* analysed in a 5–15% acrylamide gradient gel, Coomassie blue stain. The α - and β -heavy chains (α and β), the three intermediate chains (IC1, IC2 and IC3) and the light chains (LC) of the outer-arm dynein are indicated. T, tubulin. Bottom, microtubule gliding motility¹. Plus and minus refer to the ability of each fraction to support microtubule gliding motility undiluted (bottom row) or diluted with an equal volume of gradient buffer lacking DTT (top row). A 5- μ l aliquot from each gradient fraction was applied to a glass coverslip, and the protein allowed to adsorb for 5 min. Microtubules composed of purified brain tubulin and assembled using taxol were then applied with MgATP²⁻ (2 mM final concentration). Gliding movement was monitored at room temperature using video-enhanced differential interference contrast microscopy.

the substrate than is observed with MAP 1C¹. In the absence of dynein, microtubules rarely bind to the coverslip and do not glide. In the presence of dynein but without MgATP²⁻, microtubules bind to the coverslip but do not glide; binding in this

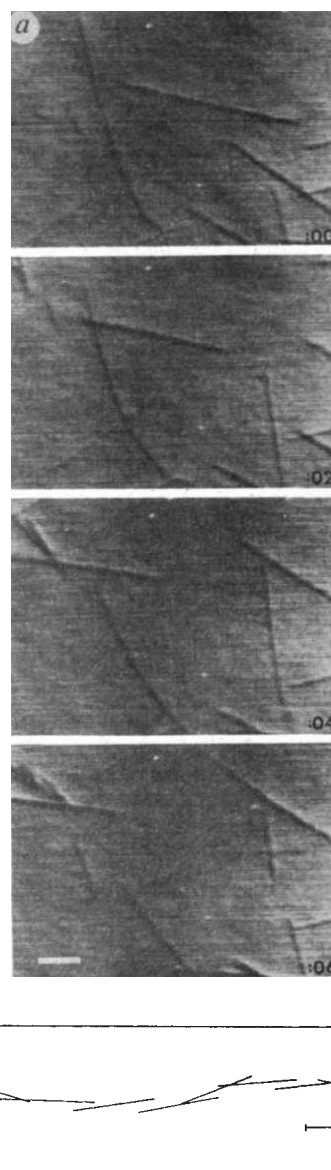


Fig. 2 Dynein-mediated microtubule gliding. Sea urchin outer-arm dynein purified by sucrose gradient centrifugation was applied to a coverslip and examined for motility as described in the legend to Fig. 1. *a*, Eight microtubules were observed to glide within this field in the 6-s interval shown. Scale bar, 1 μm . *b*, Video images of a single microtubule were traced over a 12-s interval to show the lateral mobility during translocation. Scale bar, 1 μm .

case presumably reflects the formation of rigor (ATP-free) cross-bridges.

The velocity of microtubule gliding ranges from 0.47 to $6.1 \mu\text{m s}^{-1}$ and averages $3.5 \pm 1.3 \mu\text{m s}^{-1}$ ($\bar{x} \pm \text{s.d.}$, $n = 140$) (Fig. 3). This velocity is much higher than has been reported for other microtubule translocators *in vitro*. Kinesin moves microtubules on glass with a velocity of $0.3\text{--}0.5 \mu\text{m s}^{-1}$ (ref. 17), and purified MAP 1C translocates microtubules on glass at an average velocity of $1.25 \mu\text{m s}^{-1}$ (ref. 1). In an intact sea urchin sperm flagellum beating at 45–50 Hz, the maximum rate of interdoublet sliding is $\sim 16 \mu\text{m s}^{-1}$ (ref. 18). However, the rate of doublet sliding in reactivated or disintegrating sea urchin axonemes treated to remove most of the inner or the outer arms is about half that in the intact axoneme^{7,19}. Thus, the velocities produced by purified outer-arm dynein in our *in vitro* gliding assay are comparable to those produced by a single row of arms in the axoneme.

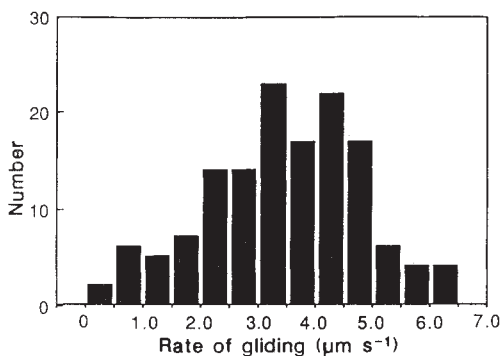


Fig. 3 Number of microtubules plotted against rate of translocation. The data are the combined total for three different experiments (140 microtubules). Measurements were made on microtubules that exhibited uninterrupted movement over an interval >0.8 s.

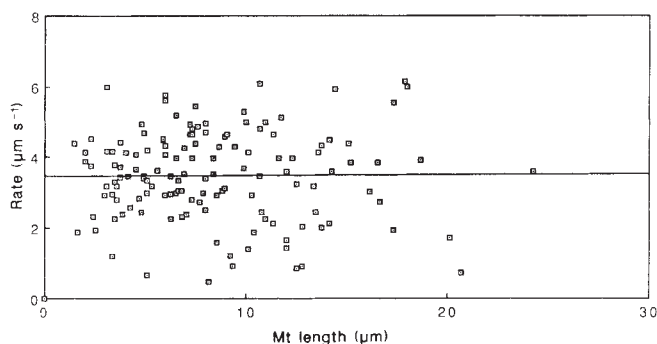


Fig. 4 Microtubule gliding rate as a function of microtubule length. The data are from the experiments shown in Fig. 3. The linear regression line ($r = 0.01$) indicates that there is no correlation between microtubule length and gliding velocity.

In studies of the disintegration of protease-treated axonemes of *Chlamydomonas*, most of the axonemes of a mutant containing outer arms but deficient in inner arms fail to undergo sliding disintegration⁸. These results raised the question of whether the outer arm can function at all in the absence of the inner arm. The findings presented here show unequivocally that the outer arm by itself is capable of microtubule translocation.

The rate of microtubule translocation is independent of microtubule length over a range of 1.5–24 μm (correlation coefficient = 0.01) (Fig. 4). Thus, the velocity of microtubule translocation is independent of the total number of arms participating in force production, as least above a critical number of crossbridges that must have been exceeded in our experiments. This is in good agreement with observations that, in disintegrating axonemes, the velocity of sliding does not depend on the extent of overlap between the sliding doublets¹⁸.

It is of interest that the velocity of microtubule gliding is much greater than predicted from the steady-state turnover rate of the purified dynein (specific ATPase activity, 0.38 μmol phosphate released per min per mg, corresponding to ~ 500 ATP min^{-1} for a dynein of 1.34×10^6 daltons; ref. 20). Assuming that one or two ATP molecules are hydrolysed per crossbridge cycle²¹ and that each cycle translocates the microtubule 4.5 nm (ref. 22), this rate of hydrolysis is sufficient to move microtubules at speeds of only 0.018–0.036 $\mu\text{m s}^{-1}$. The much greater velocities observed here could reflect the combined effect of multiple dynein arms acting on a microtubule in an asynchronous manner, and/or activation of dynein ATPase activity by the gliding microtubule. It is also possible that net translocation per ATP molecule hydrolysed is much greater than predicted from structural studies, as has been proposed for myosin²³.

The *in vitro* gliding assay described here should be useful for studying dyneins from diverse organisms, and for elucidating the functions of individual subunits of the dynein arm. This work was supported by the NIH.

Received 14 September; accepted 22 October 1987.

1. Paschal, B. M., Shpetner, H. S. & Vallee, R. B. *J. Cell Biol.* **105**, 1273–1282 (1987).
2. Paschal, B. M. & Vallee, R. B. *Nature* **330**, 181–183 (1987).
3. Summers, K. E. & Gibbons, I. R. *Proc. natn. Acad. Sci. U.S.A.* **68**, 3092–3096 (1971).
4. Sale, W. S. & Satir, P. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2045–2049 (1977).
5. Witman, G. B., Plummer, J. & Sander, G. *J. Cell Biol.* **76**, 729–747 (1978).
6. Kamimura, S. & Takahashi, K. *Nature* **293**, 566–568 (1981).
7. Yano, Y. & Miki-Nomura, T. *J. Cell Sci.* **48**, 223–239 (1981).
8. Okagaki, T. & Kamiya, R. *J. Cell Biol.* **103**, 1895–1902 (1986).
9. Sheetz, M. P. & Spudis, J. A. *Nature* **303**, 31–35 (1983).
10. Toyoshima, Y. Y. *et al. Nature* **328**, 536–539 (1987).
11. Sale, W. S., Goodenough, U. W. & Heuser, J. E. *J. Cell Biol.* **101**, 1400–1412 (1985).
12. Bradford, M. M. *Analyt. Biochem.* **72**, 248–254 (1976).
13. Pfister, K. K., Fay, R. B. & Witman, G. B. *Cell Motil.* **2**, 525–547 (1982).
14. King, S. M., Otter, T. & Witman, G. B. *Meth. Enzym.* **134**, 291–306 (1986).
15. Allen, R. D., Allen, N. S. & Travis, J. L. *Cell Motil.* **1**, 291–302 (1981).
16. Inoue, S. *J. Cell Biol.* **89**, 346–356 (1981).
17. Vale, R. D., Schnapp, B. J., Reese, T. S. & Sheetz, M. P. *Cell* **40**, 559–569 (1985).
18. Takahashi, K., Shingyoji, C. & Kamimura, S. *Symp. Soc. exp. Biol.* **35**, 159–175 (1982).
19. Gibbons, B. H. & Gibbons, I. R. *J. Cell Biol.* **13**, 337–357 (1973).
20. Gibbons, I. R. & Gibbons, B. H. in *Perspectives of Biological Energy Transduction* (eds Mukohata, Y. *et al.*) (Academic, Tokyo, in the press).
21. Shimizu, T. & Johnson, K. A. *J. biol. Chem.* **258**, 13841–13846 (1983).
22. Witman, G. B. & Minervini, N. *Symp. Soc. exp. Biol.* **35**, 203–224 (1982).
23. Yanagida, T., Arata, T. & Oosawa, F. *Nature* **316**, 366–369 (1985).

Synthesis of the skeleton of the morphine molecule by mammalian liver

Charles J. Weitz*, Kym F. Faull† & Avram Goldstein*

* Addiction Research Foundation, Palo Alto, California 94304 and Stanford University, Stanford, California 94305, USA

† Nancy Pritzker Laboratory of Behavioral Neurochemistry, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California 94305, USA

The possibility that morphine could be synthesized in animals has long been considered¹ and a pathway in mammalian brain analogous to that in the opium poppy has been proposed². Substances have been detected in mammalian brain that are recognized by antisera raised against morphine^{3–7}. Recently we reported the presence of three such immunoreactive substances in bovine hypothalamus and adrenal, and in rat brain, and the definitive identification of two of them by gas chromatography-mass spectrometry as morphine and codeine⁸. Incorporation of a labelled precursor has demonstrated the biosynthesis of morphine in the opium poppy from tyrosine-derived units⁹ (see Fig. 1). Intramolecular coupling of reticuline to form salutaridine is the critical step¹⁰ that generates the morphine skeleton (morphinan)¹¹ and the stereochemistry of the morphinan series⁹. We now report the conversion *in vivo* and *in vitro* of reticuline to salutaridine by rat liver, but this conversion is not detectable in rat brain and bovine adrenal. This is the first direct demonstration of the synthesis of a morphinan in an animal tissue and also supports the hypothesis that morphine and codeine in brain and adrenal are of endogenous origin.

To detect *in vivo* morphinan biosynthesis, we administered [³H]reticuline to rats by intracerebroventricular (250 μCi , 11.7 nmol, four experiments) or intraperitoneal (i.p.) injection (500 μCi , four experiments), removed the brains at times from 30 min to 4 h (intracerebroventricular), or the brains and livers at 1 h and 8 h (i.p.) after injection, and purified the radiolabelled products as described in the legend to Fig. 2. No peaks of radioactivity from brain extracts co-eluted in high performance liquid chromatography (HPLC) with the internal standards salutaridine, thebaine, codeine, or morphine in either type of

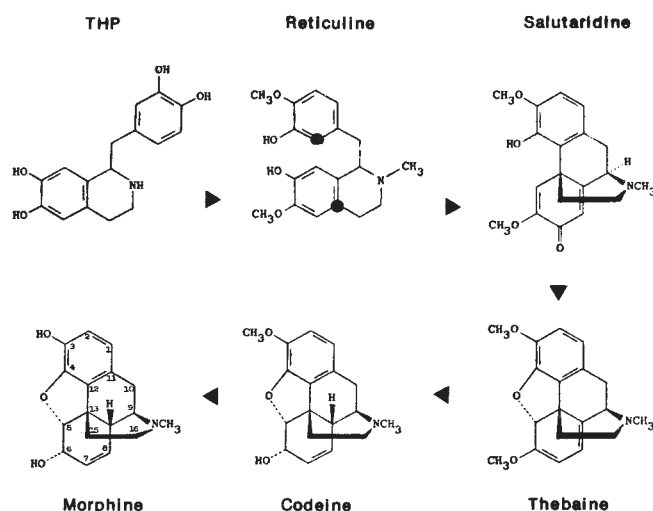


Fig. 1 Biosynthetic pathway for morphine in the opium poppy, *P. somniferum*⁹. THP (tetrahydropapaveroline, norlaudanosoline) is derived from tyrosine and dopamine. The carbon atoms marked by dots in reticuline participate in oxidative coupling to generate salutaridine, the earliest intermediate with the chiral four-ring arrangement defined as the morphinan structure.

experiment (detection limit twice background, equivalent to 8–30 fmol per brain). Figure 2a and b shows the detection of a product purified from an extract of rat liver removed 1 h after i.p. injection (a minimum of 0.92 pmol was detected); this product co-eluted with salutaridine in two HPLC systems. The adsorption of this material to the anti-morphinan immunoaffinity resin was more than 90%, whereas less than 1% was adsorbed to the non-immune control resin. In rat livers removed 8 h after injection, identical results were obtained, except that the areas of both radioactive peaks in the HPLC-M2 profile were reduced by about 75% (data not shown). No peaks of radioactivity corresponding to morphinans later in the opium poppy pathway (Fig. 1) were detected in extracts of livers removed 1 or 8 h after injection. These results were consistent with *in vivo* biosynthesis of salutaridine from reticuline in rat liver, but non-enzymatic conversion or biosynthetic conversion in some other tissue with subsequent concentration in liver could not be excluded. Furthermore, because of the large amount of radioactive precursor injected, we could not exclude the possibility of selective retention in liver (but not in brain) of a trace radioactive impurity in the [³H]reticuline stock solution.

To resolve these ambiguities we investigated a product synthesized by rat liver homogenate from [³H]reticuline. This product sequentially co-eluted with a salutaridine internal standard in two HPLC systems (Fig. 3b–d, four experiments) and was adsorbed with >90% efficiency to the immunoaffinity resin, whereas the [³H]reticuline present in the same sample was adsorbed with only 5–10% efficiency; both were adsorbed with <1% efficiency to the non-immune control resin. Although no attempt was made to optimize the conditions, we have typically seen conversion of 1% of the [³H]reticuline to this product in a 1-h incubation. Pre-boiled homogenate from the same liver gave no labelled product in the position of salutaridine (Fig. 3a; the detection limit was 1% of the peak height in the paired homogenate shown in Fig. 3b). Under identical conditions, rat brain homogenate and bovine adrenal homogenate (four experiments each) produced no detectable product with the HPLC behaviour of salutaridine (Fig. 3e, f).

To identify the biosynthesis product, we performed the experiments described in Fig. 4. By adding an excess of unlabelled reticuline to a liver homogenate containing [³H]reticuline, as described above, and purifying the labelled product, we were

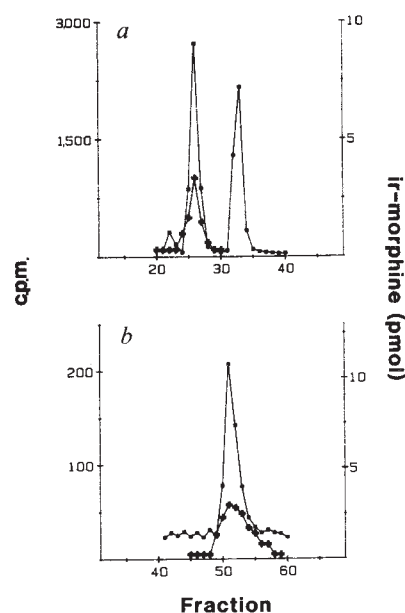


Fig. 2 Sequential co-elution in two HPLC systems of authentic salutaridine and a product purified from rat liver after i.p. administration of [³H]reticuline. *a*, HPLC-M2 (ref. 8) (Waters C₁₈, linear gradient of methanol in ammonium hydroxide; the peak of radioactivity at fraction 33 is residual [³H]reticuline). *b*, Product and salutaridine (from panel *a*) in HPLC-C2 (C₁₈, isocratic 15% acetonitrile in 5 mM trifluoroacetic acid). Closed circles (larger peaks) represent c.p.m.; crosses (smaller peaks) represent immunoreactivity with antisera against morphine (ir-morphine), as morphine equivalents. Salutaridine has 0.1% crossreactivity compared with morphine in the radioimmunoassay (RIA) (ref. 7). **Methods.** 500 μ Ci of [³H]reticuline (ring-labelled to 21.4 Ci mmol⁻¹, Amersham) in 1 ml saline was injected i.p. into a male Simonsen albino rat (300–400 g). After 55 min, the rat was anaesthetized with chloral hydrate (400 mg kg⁻¹ i.p.) and decapitated 5 min later. Brain and liver were removed, homogenized in acetic acid (0.1 M, 5 ml g⁻¹ wet weight), unlabelled carriers were added (50 pmol salutaridine, 300 pmol thebaine, 30 pmol codeine, 80 pmol morphine), and the homogenate was placed in a boiling water bath for 30 min. The sample was centrifuged, adjusted to pH 9.0, and eluted from an XAD-2 column (Amberlite; 3 ml bed volume) after washing with 35 ml of sodium carbonate buffer (10 mM, pH 9.0) (ref. 7). The eluate was dried under N₂, dissolved in sodium phosphate buffer (100 mM, pH 7.4, 18 ml), half added to anti-morphinan immunoaffinity resin (1 ml of packed, hydrated resin; antiserum 937), and the other half to non-immune resin⁷. The resins were incubated (23 °C, 4 h), filtered, washed with 9 ml of buffer, incubated in acetic acid (5 M, 9 ml) at 37 °C for 3 h, filtered, and washed with acetic acid (5 M, 9 ml). The filtrate and wash were pooled, lyophilized, dissolved in 750 μ l of 0.1 M HCl, and analysed in HPLC-C (ref. 7) (Waters C₁₈, linear gradient of acetonitrile in trifluoroacetic acid). This system resolves reticuline from thebaine, codeine, and morphine, but not from salutaridine. The peak of radioactive reticuline was collected, lyophilized, dissolved in 750 μ l of 50% methanol containing 50 nmol of salutaridine standard, and applied to HPLC-M2 (see panel *a*). Fractions were analysed by scintillation counting and RIA with antiserum 937 (ref. 7). The co-eluted peaks of radioactivity and immunoreactivity (panel *a*, fractions 24–28) were collected, lyophilized, dissolved in 750 μ l of 0.1 M HCl, and analysed in HPLC-C2 (HPLC-C modified as above; panel *b*).

able to co-purify sufficient unlabelled product for analysis by gas chromatography and mass spectrometry (GC/MS). As a positive control, a salutaridine standard was added to a parallel homogenate and subjected to all incubation and purification procedures. Figure 4b and c shows that the biosynthesis product and the salutaridine standard produced specific ion currents with retention times (14.53 and 14.52 min respectively) and

Fig. 3 Sequential co-elution in two HPLC systems of authentic salutaridine and a product synthesized from [^3H]reticuline in a rat liver homogenate. *a*, HPLC-M2 profile from a pre-boiled control liver homogenate. *b*, HPLC-M2 profile from a liver homogenate. *c*, Co-elution of the biosynthetic product (panel *b*) and salutaridine in HPLC-M2. *d*, Co-elution of the salutaridine and product (panel *c*) in HPLC-C2. *e*, HPLC-M2 profile from a rat brain homogenate. *f*, HPLC-M2 profile from a bovine adrenal homogenate. Symbols as in Fig. 2.

Methods. A fresh rat liver was homogenized in ice-cold incubation buffer (Sigma phosphate-buffered saline containing 1 mM CaCl_2 , 0.5 mM MgCl_2 , 4 ml g^{-1} wet weight), and 10 ml were removed for each incubation (equivalent to 1/5 of a liver). The control homogenate was pre-incubated (30 min) in a boiling water bath. NADPH (1 mM final), NAD (1 mM final), and [^3H]reticuline (21 μCi , 100 nM final) were added, the samples were incubated (37 $^{\circ}\text{C}$, 1 h), and termination was by addition of 8 ml glacial acetic acid (final pH 2.5). Purification as in Fig. 2 legend, except that samples were diluted to 50 ml with sodium carbonate buffer (10 mM, pH 9.0) before XAD-2 chromatography, and no salutaridine standard was added for the initial HPLC-M2. The biosynthesis peak of radioactivity (panel *b*) was collected, lyophilized, and co-chromatographed with salutaridine (50 nmol) in HPLC-M2 (panel *c*). The co-eluted peaks of radioactivity and immunoreactivity (panel *c*) were next analysed in HPLC-C2 (panel *d*). Rat brain and bovine adrenal homogenate were investigated as described above (same tissue quantity), except that in some rat brain experiments (as in panel *e*) 53 μCi (250 nM final) of [^3H]reticuline was used. Pre-boiled rat brain and bovine adrenal control homogenates (not shown) gave results identical to unboiled homogenates. Analysis of the [^3H]reticuline stock solution showed a trace impurity (0.008% abundance) eluting in the position of salutaridine in HPLC-M2; although below the detection limit in liver homogenates, it was occasionally observed in both control and experimental rat brain homogenates that had lower detection limits.

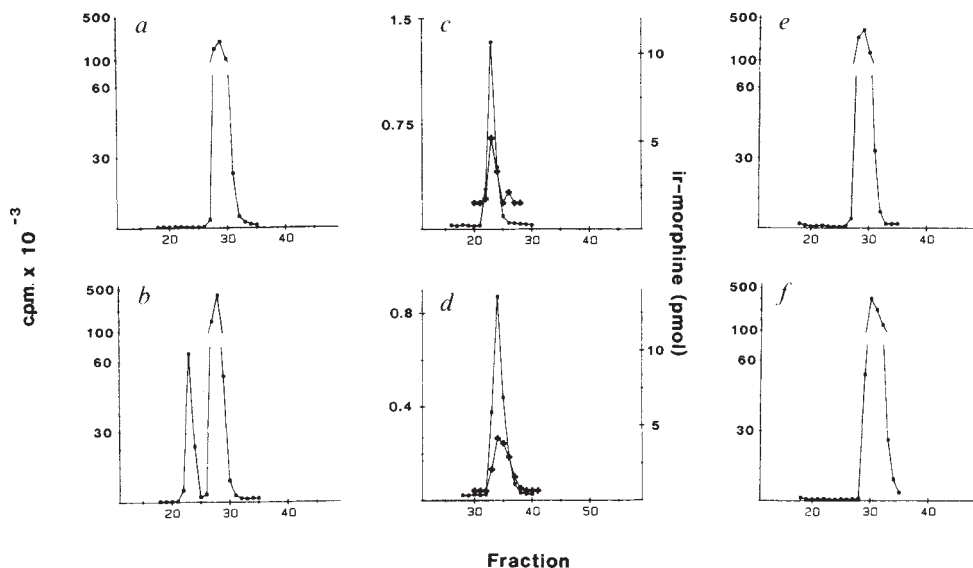
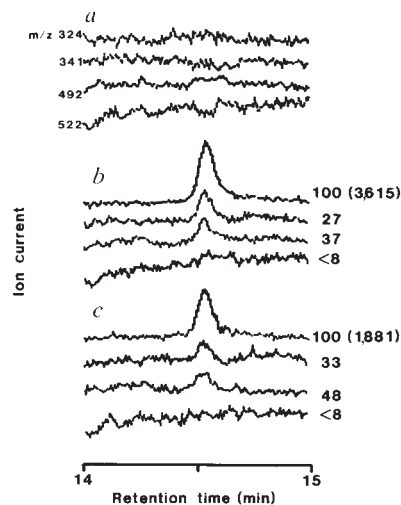


Fig. 4 Comparison in GC/MS with selected ion-monitoring of salutaridine and putative salutaridine synthesized in a liver homogenate (O-[pentafluorobenzyl]hydroxylamine derivatives): *a*, pre-boiled control; *b*, putative salutaridine biosynthesis product; *c*, salutaridine standard. The amount of salutaridine standard injected was adjusted to be equivalent to the biosynthesis product, about 20 pmol. Values of m/z for each tracing in *a* apply also to *b* and *c*. Numbers at right denote relative ion current peak areas. The absolute value of the ion current peak area for the $m/z = 324$ tracing is in parentheses.

Methods. Three 10-ml homogenate samples were prepared from a single rat liver; the pre-boiled control and experimental homogenates were as described (Fig. 3 legend), except that unlabelled reticuline (7.1 nmol) was added at the start of incubation in addition to [^3H]reticuline. To the third homogenate, a salutaridine standard (500 pmol) was added at the start of incubation, but no unlabelled reticuline was added. The homogenates were processed through the initial HPLC-M2 (Fig. 3 legend), except that unlabelled reticuline (7.1 nmol) was added to the salutaridine standard homogenate just before termination, no other carriers were added, and the non-immune resins were omitted. Samples were eluted from HPLC-C2, pooled, lyophilized, dissolved in methanol, transferred to glass tubes, and dried in a vacuum concentrator. O-(Pentafluorobenzyl)hydroxylamine-HCl (20 μl ; Pierce) was added and the tubes were sealed by flame and incubated (75 $^{\circ}\text{C}$, 1 h). Water (10 μl) was added, the samples extracted three times with heptane (10 μl), and the pooled heptane phases dried under N_2 . Samples were dissolved in heptane (5 μl) for GC/MS analysis as described⁸ with the following modifications: electron capture-chemical ionization (methane, 1 torr) with negative ion detection; selected ion monitoring; column length 15 m; 100 $^{\circ}\text{C}$ for 1 min, then 20 $^{\circ}\text{C} \text{ min}^{-1}$ to a plateau of 300 $^{\circ}\text{C}$. Detection of a peak for the molecular ion ($m/z = 522$) would not be expected with the sample quantities above.



relative peak areas that were identical within experimental error. The quantity of biosynthetic salutaridine recovered after purification, estimated from the known specific activity of the reticuline, was 18 pmol, and that estimated from GC/MS ion current measurements in comparison with a salutaridine standard was 22 pmol (not shown). A pre-boiled control homogenate produced no labelled product at the elution position of salutaridine (detection limit 1% of the peak height in the salutaridine position in the unboiled homogenate; result identical to that in Fig. 3*a*), nor did it generate selected ion currents with the retention time of derivatized salutaridine (Fig. 4*a*).

These results demonstrate that mammalian liver can carry out the critical step in morphine biosynthesis by converting reticu-

line to salutaridine. Neither reticuline nor salutaridine has yet been identified in animal tissue, but the unusual features of this intramolecular coupling reaction suggest that it is catalysed by a specific enzyme. The existence of morphinan synthesis activity in mammalian liver suggests that the morphine and codeine present in mammalian brain and adrenal³⁻⁸ are of endogenous origin. The lack of activity in brain and adrenal homogenates suggests that one or more steps in any potential mammalian biosynthesis of morphine takes place outside these tissues. Whether a complete pathway for morphine biosynthesis exists in animals needs to be further investigated.

This work was supported by a grant to A.G. from the National Institute on Drug Abuse. K.F.F. was supported in part by grants

to Dr J. D. Barchas from the National Institute of Mental Health and the National Institute on Drug Abuse. We thank Dr E. Brochmann-Hanssen for reticuline, Dr H. Rapoport for salutaridine, Dr Gottfried Feistner for advice regarding mass spectrometry, Dr A. Roy for providing fresh tissue, and Louise I. Lowney for helpful comments.

Received 31 July; accepted 13 October 1987.

1. Mavroujanis, M. *Comptes rendue Soc. Biol.* **55**, 1092-1094 (1903).
2. Davis, V. E. & Walsh, M. J. *Science* **167**, 1005-1007 (1970).
3. Gintzler, A. R., Levy, A. & Spector, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2132-2136 (1976).
4. Gintzler, A. R., Gershon, M. D. & Spector, S. *Science* **199**, 447-448 (1978).
5. Killian, A. K. *et al. Life Sci.* **28**, 811-817 (1981).
6. Donnerer, J., Oka, K., Brossi, A., Rice, K. C. & Spector, S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4566-4567 (1986).
7. Goldstein, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5203-5207 (1985).
8. Weitz, C. J., Lowney, L. I., Faull, K. F., Feistner, G. & Goldstein, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9784-9788 (1986).
9. Kirby, G. W. *Science* **155**, 170-173 (1967).
10. Hodges, C. H. & Rapoport, H. *Biochemistry* **21**, 3729-3734 (1982).
11. Barton, D. H. R. *et al. J. Chem. Soc. C*, 2423-2438 (1965).

A new petunia flower colour generated by transformation of a mutant with a maize gene

Peter Meyer*, Iris Heidmann*, Gert Forkmann† & Heinz Saedler*

* Max-Planck-Institut für Züchtungsforschung, D-5000 Köln 30, FRG
† Universität Tübingen, Institut für Biologie II, Lehrstuhl für Genetik, D-7400 Tübingen, FRG

Petunia hybrida is one of the classical subjects of investigation in plants in which the pathway of anthocyanin biosynthesis has been analysed genetically and biochemically. In *petunia* cyanidin- and delphinidin-derivatives, but no pelargonidin-derivatives are produced as pigments. This is due to the substrate specificity of the dihydroflavonol 4-reductase of *petunia*, which cannot reduce dihydrokaempferol. The *petunia* mutant RL01, which accumulates dihydrokaempferol, shows no flower pigmentation. RL01 served as a recipient for the transfer of the *A₁* gene of *Zea mays* encoding dihydroquercetin 4-reductase, which can reduce dihydrokaempferol and thereby provided the intermediate for pelargonidin biosynthesis. Transformation of RL01 with a vector p35A1, containing the *A₁*-complementary DNA behind the 35S promoter leads to red flowers of the pelargonidin-type. Thus a new flower pigmentation pathway has been established in these plants.

Dihydroflavonol 4-reductase (DFR) is specifically involved in anthocyanin biosynthesis. In general, the enzyme catalyses the conversion of the dihydroflavonols dihydrokaempferol, dihydroquercetin and dihydromyricetin to the flavan-3,4-cis-diols leucopelargonidin, leucocyanidin and leucodelphinidin, respectively, which are the immediate precursors for the respective anthocyanidins¹. The *petunia* DFR, however, does not accept dihydrokaempferol as substrate, but only reduces dihydromyricetin and, less efficiently, dihydroquercetin². This special substrate specificity probably explains the virtual absence of anthocyanins based on pelargonidin from *Petunia hybrida*³. For this reason *petunia* mutants, which do not contain the appropriate precursors dihydroquercetin or dihydromyricetin, should exhibit a white flower phenotype. The mutant RL01 from the *Petunia hybrida* collection at Tübingen is a derivative of line R4⁴ which has previously been used for genetic and enzymatic studies on B-ring hydroxylation^{4,5} and flavonol formation⁶. Because RL01 contains recessive alleles at the *Ht1* and *Hf1* loci, it lacks flavanoid 3'-hydroxylase activity and exhibits only minor flavonoid 3',5'-hydroxylase activity⁴ resulting in the accumulation of significantly reduced amounts of cyanidin and delphinidin derivatives and hence to a pale pink flower

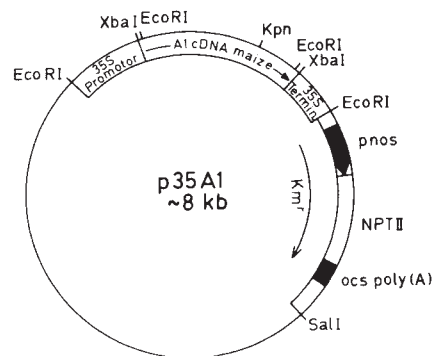


Fig. 1 Construction of plasmid p35A1. In a 1,320 base pair (bp) *EcoRI* fragment of a full size cDNA clone of a type 2 *A₁* gene of *Zea mays*⁸ the *EcoRI* restriction sites were filled in and *XbaI*-linkers were attached which restored the filled in *EcoRI* sites. The *XbaI* fragment was cloned into the unique *XbaI* site of plasmid pCKan1, where it is located between 35S promoter and terminator sequence of CaMV. The large *EcoRI* fragment derives from plasmid pLGV11, which allows selection of transformants by kanamycin. Plasmid pLGV11 is equivalent to pLGV1103¹² except for the deleted Tn903 *SalI* fragment.

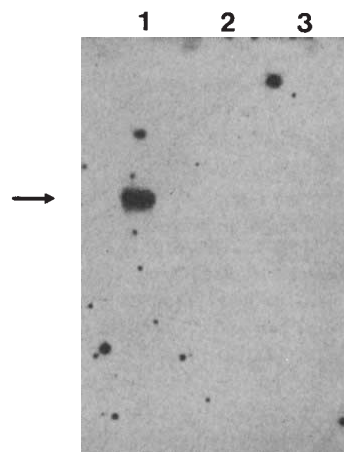


Fig. 2 Transcription of the *A₁* gene in transgenic plants. Plant RP235-15, which exhibits the new pelargonidin pigment, shows strong messenger RNA expression of the maize *A₁* gene (lane 1) whereas no *A₁* transcript is detectable in transformant RP235-12 with an unchanged flower colour (lane 2) and in RL01 (lane 3). **Method.** mRNA was extracted from leaves as described¹⁶, using Hybond™-mAP (Amersham) according to the manufacturers instructions. Hybridization was done with an oligomer-primed probe^{17,18}.

phenotype. Anthocyanins based on pelargonidin are not formed in this mutant. Instead the respective precursor dihydrokaempferol is accumulated due to the mutations mentioned above (Fig. 4). If leucopelargonidin is supplemented to RL01, pelargonidin pigments are produced (G. F., unpublished data), suggesting that all other genes necessary for anthocyanin biosynthesis⁷ are represented by dominant alleles. RL01 is apparently only defective in the hydroxylation of dihydrokaempferol.

Protoplasts of this line were transformed with DNA of plasmid p35A1 (Fig. 1) which contained a cDNA clone of the *A₁* gene from *Zea mays*⁸. The *A₁* gene encodes dihydroquercetin 4-reductase (DQR) which converts dihydroquercetin into leucocyanidin leading to the production of cyanidin derivatives in the maize aleurone⁹. The coding sequence of the *A₁* gene was inserted between the promoter and the terminator of the 35S gene from CaMV and cloned into the plant expression



Fig. 3 RL01 (left) exhibits a pale pink flower phenotype due to traces of cyanidin and delphinidin derivatives. Transformant 235-15 (right) shows brick red pelargonidin coloration. **Methods.** After M-phase transformation¹⁰ microcalli were cultured and selected in bead type culture¹³ in V47-medium¹⁴ reducing osmolarity by 100 mosm per week. Kanamycin resistant microcalli were transferred to regeneration medium when they reached a diameter of 3–5 mm. They were placed on Re27/6-medium molar strength (MS)-medium¹⁵ with 2mg per 1 benzylaminopurine (BAP) and 2 mg per 1 indoleacetic acid (IAA) for three weeks and transferred to Re17/3 (MS-medium with 1 mg per 1 BAP and 1 mg per 1 IAA). Shoots were rooted on MS-medium without hormones. All media contained kanamycin (50 mg per l).

vector pLGV11 neo (Fig. 1). The resulting plasmid (p35A1) was transferred into protoplasts of RL01 which had been synchronized into M-phase before transformation¹⁰. Two per cent of the surviving calli expressed the plasmid-borne *NPT II* gene and were consequently resistant to kanamycin. From each transformed callus two plants were regenerated.

Of the first 15 flowering transformants, two transformed plants showed a brick red coloration on the flowers of both regenerated plants (Fig. 3), another four had flowers with brick red sectors of differing sizes. The latter phenotype is not fully understood and will be subject of further investigations. One transformant (RP235-15), all the flowers of which were uniformly brick red in colour and in which transcription of the *A₁* gene was clearly detectable (Fig. 2), was used for flavonoid analysis and compared to RL01 using standard procedures¹¹. In addition to cyanidin-3-glycoside, cyanidin-3-glycosylglycoside and delphinidin-3-glycoside, which are weakly expressed in both RL01 and RP235-15, RP235-15 contains pelargonidin-3-glycoside and pelargonidin-3-glycosylglycoside as major components. The spectrophotometric peak of anthocyanin was shifted from 528 nm in RL01 to 512 nm in RP235-15. Furthermore no dihydrokaempferol and only traces of kaempferol, both of which are accumulated in RL01, are detectable in RP235-15.

This analysis strongly suggests that the product of the introduced maize *A₁* gene created a pathway new to *Petunia hybrida*.

Our data demonstrate that the DQR of maize does not show the same narrow substrate specificity as the petunia reductase, but converts dihydrokaempferol into leucopelargonidin which is then processed further to pelargonidin-3-glycosides, leading to the red flower pigment (Fig. 4). The p35A1 vector in conjunction with the RL01 petunia mutant incorporates the *A₁* gene, a

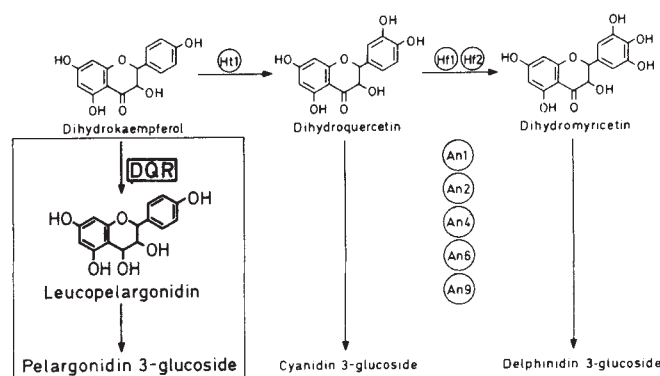


Fig. 4 Pathway diagram. Due to mutations in flavonoid 3'-hydroxylase (Ht1) and flavonoid 3',5'-hydroxylase (Hf1, Hf2) dihydrokaempferol is only poorly hydroxylated to dihydroquercetin and dihydromyricetin in RL01. Dihydrokaempferol is taken as a substrate by maize DQR and converted into leucopelargonidin which is further processed into pelargonidin-3-glycosides leading to the brick red colour.

marker with an easily visible phenotype which will be useful for many purposes, including the study of flower specific expression signals.

We thank Drs P. Starlinger, K. Hahlbrock, A. Pereira and D. Lydiat for critical reading of the manuscript, Dr Z. Schwarz-Sommer for the cDNA clone of *A₁*, Dr A. Pereira for pCKan1, I. Niedenhof, T. Akmandor and A. Pinten for technical assistance and F. Furkert, M. Kalda and D. Bock for illustrations and photographs.

Received 1 September; accepted 27 October 1987.

- Heller, W., Forkmann, G., Britsch, L. & Grisebach, H. *Planta* **165**, 284–287 (1985).
- Forkmann, G. & Ruhnau, B. *Z. Naturforsch.* **42c**, 1146–1148 (1987).
- Cornu, A., Paynot, M. & Touvin, H. *Phytochemistry* **13**, 2022 (1974).
- Stotz, G., de Vlamming, P., Wiering, H., Schramm, A. W. & Forkmann, G. *Theor. appl. Genet.* **70**, 300–305 (1985).
- Stotz, G. thesis, Univ. Tübingen (1983).
- Forkmann, G., de Vlamming, P., Spribille, R. & Schram, A. W. *Z. Naturforsch.* **41c**, 179–186 (1986).
- Wiering, H. & de Vlamming, P. in *Petunia Monograph* (ed. Sink) 49–67 (Springer, Berlin, 1984).
- Schwarz-Sommer, Z. *et al. EMBO J.* **2**, 287–294 (1987).

- Reddy, A. R., Britsch, L., Salamini, F., Saedler, H. & Rohde, W. *Pl. Sci.* **52**, 7–12 (1987).
- Meyer, P., Walgenbach, E., Bussmann, K., Hombrecht, G. & Saedler, H. *Molec. gen. Genet.* **201**, 513–518 (1985).
- Harborne, J. B. *Comparative Biochemistry of the Flavonoids* (Academic, London & New York, 1967).
- Hain, R. *et al. Molec. gen. Genet.* **199**, 161–168 (1985).
- Shilito, R. D., Paszkowski, I. & Potrykus, I. *Pl. Cell Rep.* **2**, 244–247 (1983).
- Binding, H. *Z. Pflanzenphysiol.* **74**, 327–356 (1974).
- Murashige, T., Skoog, F. *Physiol. Pl.* **15**, 473–497 (1962).
- Logemann, J., Schell, J. & Wilmshier, L. *Analyt. Biochem.* **163**, 16–20 (1987).
- Feinberg, A. P. & Vogelstein, R. *Analyt. Biochem.* **132**, 6–13 (1983).
- Feinberg, A. P. & Vogelstein, R. *Analyt. Biochem.* **137**, 266–267 (1987).